Feedback suppression in digital hearing instruments

Ma, Guilin

Publication date:
2010

Technical University
of Denmark

Feedback Suppression in Digital Hearing Instruments

Guilin Ma

April 2010

DTU Electrical Engineering
Department of Electrical Engineering

Technical University of Denmark
Department of Electrical Engineering
Acoustic Technology
DK-2800 Kgs. Lyngby
Denmark

GN ReSound A/S
Lautrupbjerg 7
DK-2750 Ballerup
Denmark

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Technical University of Denmark.

Preface

This dissertation is the documentation of my Ph.D. work at Acoustic Technology, Electrical Engineering, Technical University of Denmark (DTU) and GN ReSound A/S from May 1st 2007 to April 30 2010. It serves as a partial fulfillment of the requirements for the degree of Doctor of Philosophy.

The topic of this dissertation is feedback suppression in digital hearing instruments. The project is under the supervision of Finn Jacobsen at DTU (main supervisor), Finn T.Agerkvist at DTU (co-supervisor), Jim Luther at GN ReSound A/S for the first year and Fredrik Gran at GN ReSound A/S for the remaining two years.

The published/submitted papers during my Ph.D. study are attached at the end of the dissertation.

Guilin Ma
Kgs. Lyngby, April 2010

Acknowledgments

I would like to express my sincere acknowledgements to all my supervisors for their suggestions, support and encouragement during the project, especially to Fredrik Gran and Finn Jacobsen. Dr. Gran has helped me pull through the difficult beginning of the project, led me in the right direction towards a researcher and provided valuable inputs to my project. Dr. Jacobsen has been very patient whenever I have problems with the study and he has helped me improve my writing skills constantly.

I am also grateful to all the colleagues at GN Resound A/S. They are very friendly and helpful.

Finally, I would like to thank my wife Min Wen, who has supported me all along. I also would like to thank my parents. It is they who have brought me up and given me endless love.

Abstract

In this dissertation, the feedback whistling problem with digital hearing instruments is investigated. The work focuses on the properties of the feedback path, the modelling of the feedback path and the feedback suppression techniques.

The properties and modelling of the feedback path are first discussed. Along the propagation path, different components compose the feedback path. The effects of these components are analyzed and categorized. Accordingly an ideal feedback path model, which consists of a fixed model, a slowing varying model and a fast varying model, is suggested in the dissertation. Methods to extract the fixed model are proposed and proved to be effective in representing the invariant part of the feedback path. Based on the investigation of the dynamic changes of the feedback path in adverse situations, for example when the user picks up the telephone handset, a reflection model is developed as one type of the fast varying models.

The techniques to suppress the feedback are then reviewed. To improve the existing feedback suppression systems, two approaches are proposed to address the so-called "bias problem". The first approach improves the performance of the adaptive feedback canceller with filtered-X adaptation by injecting nearly inaudible noise. The second approach uses a linear predicative coding based vocoder to synthesize the hearing-aid output in order to decorrelate the hearing-aid output signal and the desired input signal.

In the end, a discussion about the use of the proposed feedback path models in the feedback cancellation systems is presented.

Contents

Preface iii

Acknowledgments v

Abstract vii

Nomenclature xiii

1 Introduction 1
- 1.1 Problem Formulation . . . 2
- 1.2 Contributions . . . 2
- 1.3 Structure of the Dissertation . . . 4

2 Digital Hearing Aids 5
- 2.1 History of Hearing Aids . . . 6
- 2.2 Digital Signal Processing in Hearing Aids . . . 7
- 2.3 Problems with Digital Hearing Aids . . . 8

3 Feedback Path Modelling 11
- 3.1 Feedback Whistling . . . 11
- 3.2 Properties of the Feedback Path . . . 13
  - 3.2.1 Components of the Feedback Path . . . 13
  - 3.2.2 Properties of the Dynamic Feedback Path . . . 15
- 3.3 Modelling of the Feedback Path . . . 18

4 Feedback Suppression Techniques 21
- 4.1 Forward-path Suppression . . . 21
- 4.2 Adaptive Feedback Cancellation . . . 22
- 4.3 Bias of the Adaptive Feedback Cancellation . . . 24
- 4.4 Bias Reduction Techniques . . . 25
- 4.5 Feedback Path Models used in the Adaptive Feedback Cancellation . . . 28

5 Conclusions and Suggestions for Further Research 31

5.1 Conclusions 31
5.2 Suggestions for Future Research 32

Paper I **35**
Using a Reflection Model for Modeling the Dynamic Feedback Path of Digital Hearing Aids
G. Ma, F. Gran, F. Jacobsen and F. T. Agerkvist
Journal of the Acoustical Society of America, Vol. 127, No. 3, pp. 1458-1468, March 2010

Paper II **49**
Adaptive Feedback Cancellation with Band-limited LPC Vocoder in Digital Hearing Aids
G. Ma, F. Gran, F. Jacobsen and F. T. Agerkvist
Reviewed, revised and resubmitted to IEEE Transactions on Audio, Speech and Language Processing, March 2010

Paper III **63**
Extracting the Invariant Model from the Feedback Paths of Digital Hearing Aids
G. Ma, F. Gran, F. Jacobsen and F. T. Agerkvist
Submitted to Journal of the Acoustical Society of America, April 2010

Paper IV **77**
Noise Injection for Feedback Cancellation with Linear Prediction
G. Ma, F. Gran, F. Jacobsen and F. T. Agerkvist
Accepted as a Contributed Paper by the International Congress on Acoustics, 2010

Paper V **83**
A New Approach for Modelling the Dynamic Feedback Path of Digital Hearing Aids
G. Ma, F. Gran, F. Jacobsen and F. T. Agerkvist
Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, April 2009

Bibliography **89**

A Additional Publication **93**

Paper VI **95**
Monaural Separation of Dependent Audio Sources Based on a Generalized Wiener Filter
G. Ma, F. T. Agerkvist and J. B. Luther

Proceedings of the IEEE International Symposium on Signal Processing and Information Technology, December 2007

Nomenclature

$\mathcal{Z}\{\cdot\}$	Z-transform
$w(n)$	Time series with time index n
W	Z-transform of the time series w(n)
$\mathbf{w}$	Vector (bold notation)
$\mathbf{W}$	Matrix (bold notation, capital letter)
$\{\cdot\}^T$	Matrix transpose (non-conjugate)
$\{\cdot\}^{-1}$	Matrix inverse

Abbreviations

AFC	Adaptive Feedback Canceller
AR	Autoregressive
BTE	Behind-The-Ear
BLPC	Band-limited Linear Predictive Coding
BLPC-AFC	Band-limited Linear Predictive Coding based Adaptive Feedback Canceller
BLPC-FxAFC	Band-limited Linear Predictive Coding and Filtered-X adaptation based Adaptive Feedback Canceller
DSP	Digital Signal Processing
FIR	Finite-Impulse-Response
IIR	Infinite-Impulse-Response
ITC	In-The-Canal
ITE	In-The-Ear
LPC	Linear Predictive Coding
LMS	Least Mean Squares
MSE	Mean Square Error
MSG	Maximum Stable Gain
NLMS	Normalized Least Mean Squares
PEM-AFC	Prediction Error Method based Adaptive Feedback Canceller
RLS	Recursive Least Squares

CHAPTER

ONE

Introduction

The demographics of hearing loss show that an estimated 500 million people experience hearing loss worldwide today [1], which amounts to 7% of the total population. The hearing aid is a medical device that partially overcomes the deficits associated with a hearing loss [2].

Feedback whistling is a severe problem with hearing aids. It limits the maximum stable gain (MSG), which is the maximum amplification the hearing aids can provide, in most hearing aids [3] and compromises the comfort of wearing. In modern hearing aids, feedback suppression has become an indispensable function to provide enough amplification for the users and to improve the comfort. After years of research, the feedback suppression algorithm in top-segment hearing aids today has been developed into so sophisticated a scheme that it can prevent the feedback whistling in normal situations and even in some adverse situations.

However, the feedback whistling is still in the areas with the highest negative ratings in the latest survey of customer satisfaction of hearing aids [4]. Therefore, research on the improvement of the feedback suppression is still of high importance from user point of view.

This work serves as such an endeavor of improving the existing feedback suppression systems. The major focus is on the feedback path modelling and the techniques used in the adaptive feedback cancellation.

1.1 Problem Formulation

The presence of feedback has long been one of the major problems in the fitting and wearing of hearing aids. It occurs when a part of the amplified sound leaks from the ear canal and is picked up by the hearing-aid microphone, re-amplified and sent out by the hearing-aid receiver again. Unchecked feedback can lead to system instability, which limits the MSG that can be achieved in most hearing aids, and brings about very unpleasant whistling.

No matter how carefully the hearing aid and the earmold are fitted, the feedback problem may still occur in most hearing aids because they are vented for reasons of comfort [5]. The feedback problem is especially serious when the wearers chew, talk, put on a hat, pick up a phone or comb their hair. As the new open-fitting concept becomes more and more appealing in the hearing-aid market today, the improvement of feedback suppression is even more desirable because the large opening exacerbates the feedback problem.

The feedback suppression has been investigated for decades, and various approaches have been proposed. However, in practice the performance is limited either by the inherent problems of the algorithms or by the diversity of external acoustical environments.

The objective of this project is twofold: Firstly, it aims at solving the inherent problems of the feedback suppression algorithms; secondly, it tries to improve the performance of the feedback suppression in the most adverse acoustical environments.

1.2 Contributions

This work has mainly investigated two important issues:

- **Feedback path modelling**
- **Bias reduction in the adaptive feedback cancellation**

For the first issue, an ideal feedback path model that consists of a fixed model, a slowly varying model and a fast varying model is suggested and two new models are proposed. One serves as one type of the fast varying models, which could potentially be used in the feedback cancellation under the most

adverse conditions. The other attacks the extraction of the fixed model. These two models have been elaborated in three papers:

- **G. Ma**, F. Gran, F. Jacobsen and F. T. Agerkvist, "Using a Reflection Model for Modeling the Dynamic Feedback Path of Digital Hearing Aids", *Journal of the Acoustical Society of America*, Vol. 127, No. 3, pp. 1458-1468, March 2010.
- **G. Ma**, F. Gran, F. Jacobsen and F. T. Agerkvist, "Extracting the Invariant Model from the Feedback Paths of Digital Hearing Aids", *Submitted to Journal of the Acoustical Society of America*, April 2010.
- **G. Ma**, F. Gran, F. Jacobsen and F. T. Agerkvist, "A New Approach for Modelling the Dynamic Feedback Path of Digital Hearing Aids", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, April 2009.

During the investigation of the second issue, which is related to the primary problem of the adaptive feedback canceller (AFC), several methods have been proposed. This work has led to several papers and several patent applications:

- **G. Ma**, F. Gran, F. Jacobsen and F. T. Agerkvist, "Adaptive Feedback Cancellation with Band-limited LPC Vocoder in Digital Hearing Aids", *Reviewed, revised and resubmitted to IEEE Transactions on Audio, Speech and Language Processing*, March, 2010.
- **G. Ma**, F. Gran, F. Jacobsen and F. T. Agerkvist, "Noise injection for feedback cancellation with linear prediction", *Accepted as a Contributed Paper by the International Congress on Acoustics*, Sydney, 2010.
- **G. Ma**, F. T. Agerkvist and J. B. Luther, "Monaural Separation of Dependent Audio Sources Based on a Generalized Wiener Filter", *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology*, Cairo, Egypt, December 2007.
- **G. Ma**, F. Gran, "Hearing Aid with Means for Decorrelating Input and Output Signals", US Patent Application 12/580,864, October 16, 2009, and EP Patent Application 09170198.7, September 14, 2009.
- F. Gran, **G. Ma**, "Hearing Aid with Means for Adaptive Feedback Compensation", US Patent Application 12/580,888, October 16, 2009, and EP Patent Application 09170200.1, September 14, 2009.

1.3 Structure of the Dissertation

The dissertation is divided into the following chapters.

Chapter 2 provides an introduction to the digital hearing aids including the history, the benefits and the problems.

Chapter 3 discusses the feedback path modelling. This chapter starts with the properties of feedback path, then introduces the existing feedback path models and at the end describes the two new models proposed in this work.

Chapter 4 presents the techniques of feedback suppression. This chapter reviews the various approaches that have been proposed, explains the so-called "bias problem" in the adaptive feedback cancellation, introduces different techniques for bias reduction, describes two methods proposed in this work and at the end links the feedback path modelling with the adaptive feedback cancellation.

Chapter 5 concludes this dissertation and provides suggestions for future work.

Paper I-V are included as part of the framework in Chapter 3 and Chapter 4.

Appendix A includes an additional paper that is not covered in this dissertation.

CHAPTER

TWO

Digital Hearing Aids

The anatomy of a human ear is shown in Fig. 2.1. The outer ear consists of pinna and ear canal. The middle ear consists of ear drum, malleus, incus and stapes. The inner ear consists of cochlea and nerves.

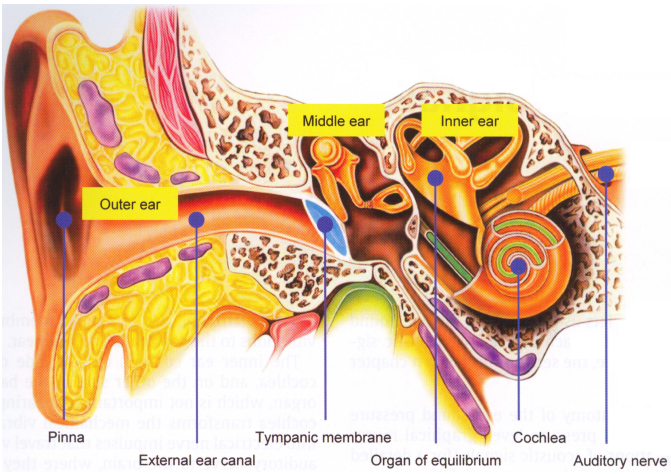


Figure 2.1: The human ear. (From [6])

Hearing loss, caused by damage in different parts of the ear, falls into two types: conductive hearing loss and perceptive hearing loss. The first type is usually attributed to problems in the middle ear. The second is resulted from cochlear problems (in the inner ear) caused by age and/or noise, or retro-cochlea problems (after inner ear, in the nerve system from cochlea to the brain). The vast majority of perceptive hearing loss is associated with abnormalities in the hair cells in the cochlea.

Hearing loss is not only a reduced sensitivity shown as an elevated threshold in the audiogram, but also a problem associated with abnormal loudness growth (recruitment), reduced frequency selectivity (broader critical bands and more masking), reduced temporal resolution and reduced binaural processing. Thus, the hearing aid is far more than an amplifier; it is a device that can help the hearing-impaired person both detect and make effective use of acoustical signals [7].

2.1 History of Hearing Aids

The history of hearing aids can be traced back to the 19th century. They began as large trumpets and horns as shown in Fig. 2.2.

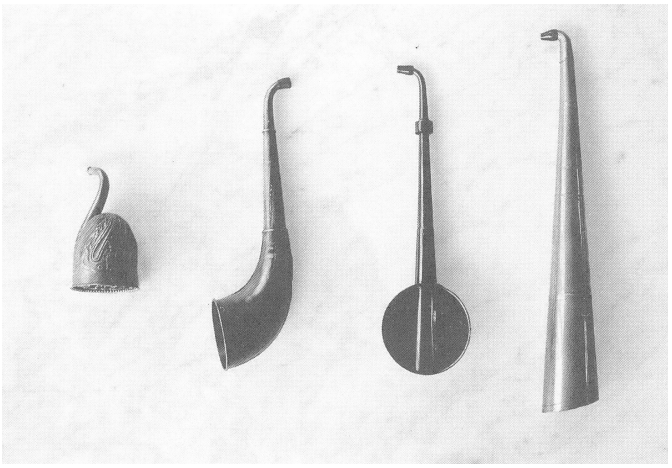


Figure 2.2: Various types of ear trumpets. (From [7])

Later, they evolved into carbon hearing aids, vacuum-tube hearing aids and transistor hearing aids. These devices were still bulky and uncomfortable to wear. During the mid-1980s the first digital hearing aid was launched. It was not until ten years later that digital hearing aids really became successful, with small digital devices placed either inside or discreetly behind the ear.

The modern hearing aids not only amplify sounds, but also process the sounds so that they can be understood by the users in noisy conditions. In modern hearing aids many auxiliary functions are incorporated, for example, datalogging and connection with accessories. A non-exclusive list of modern hearing-aid functions is:

- Compression
- Single-channel Noise Reduction
- Feedback Cancellation
- Directionality/Beamforming
- Wind Noise Reduction
- Sound Classification
- Datalogging

A typical block diagram of digital hearing aids is show in Fig. 2.3. The lower part of the diagram shows the signal path through the hearing aid from microphone to receiver. The upper three blocks are auxiliary control functions.

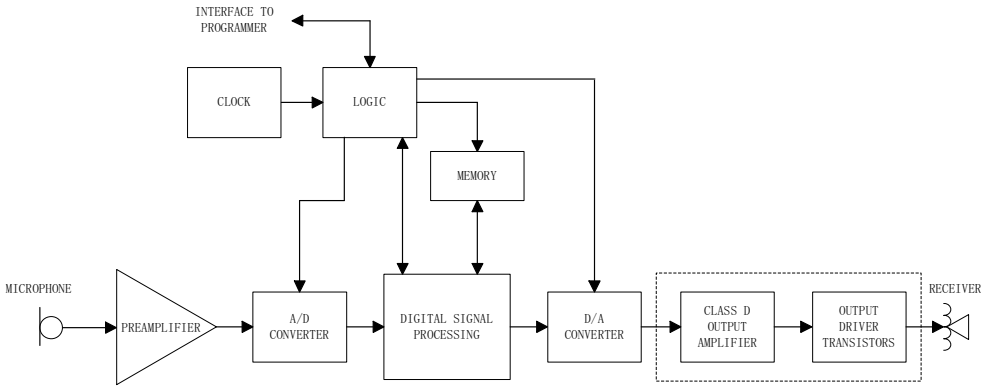


Figure 2.3: Generic block diagram of a DSP based hearing aid. (From [8])

2.2 Digital Signal Processing in Hearing Aids

The introduction of a survey [9] states that digital hearing aids offer eight significant advantages that were not available before with analog hearing aids:

- Superior signal processing capabilities, increasing the chances that noise sources will be removed and that the instrument will capture and understand more of the speech signal, or that some sounds will be enhanced to aid speech intelligibility.

- Active noise reduction and cancellation and therefore greater user comfort in noisy situations.
- Greater flexibility in fitting the instrument to the unique hearing loss characteristics of the consumer.
- Better ability to reduce internal noise in the hearing instrument through suppression of acoustical and mechanical feedback.
- Superior optimization of microphones in directional hearing instruments.
- Better overall shaping of the frequency response.
- The ability, through datalogging, to use DSP to better monitor hearing instrument use, which will aid the fine-tuning process for some consumers.
- Overall cleaner sound delivered to the consumer's ears.

Most of these advantages are attributed to the signal processing used in the digital hearing aids, which is carried out by one or more processors. As the technology of integrated circuits rapidly advances, the hearing aids can provide more and more algorithms to meet the users' needs.

However, the complexity of the algorithms in the hearing aid is still limited by the amount of power that can be supplied by the small battery. When the wireless technology is introduced to the hearing aids, the power consumption issue is even more severe. Therefore, the research on the algorithms of the hearing aids is subject to the constraint of complexity.

2.3 Problems with Digital Hearing Aids

Modern hearing aids are proved to be helpful to hearing-impaired people. However, no more than 40% of those people over 65 actually use hearing aids. Besides the cost factor, comfort is also an important issue.

The discomfort may occur in many situations. The wearers could feel the sounds unnatural. They might encounter localization difficulties in daily use. They could hear their own voice very loud due to the occlusion effect. They often experience loud whistles when picking up the phone.

Among the problems with digital hearing aids, the feedback whistling is ranked as No.3 of the hearing-aid improvements sought by current hearing aid owners [10]. This work focuses on the feedback whistling problem of the hearing aids.

CHAPTER

THREE

Feedback Path Modelling

To address the feedback whistling problem, one has to understand two problems: what is the feedback problem? what are the properties of the feedback path? Based on the answers to these questions, the feedback path can be modelled.

3.1 Feedback Whistling

The term feedback literally means that some of the output of the hearing aids manages to get back to the input, i.e., it is fed back to the input [2]. In hearing aids, two kinds of feedback paths exist: mechanical feedback and acoustical feedback. The mechanical feedback is usually caused by the vibrations of the receiver diaphragm being transmitted back to the microphone diaphragm through contact with the hearing-aid shell. The acoustical feedback shown in Fig. 3.1 includes the effects of the hearing-aid amplifier, receiver, microphone, the acoustics of the vent or leak as well as the external acoustics.

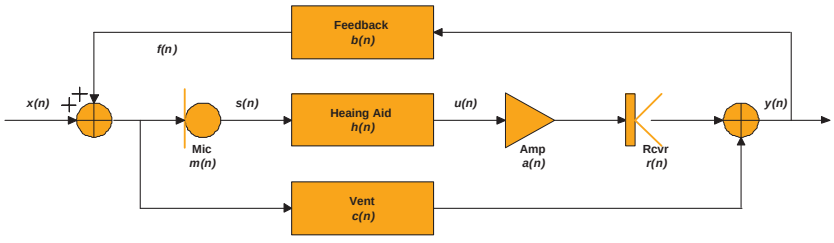


Figure 3.1: Block diagram of a generic hearing aid without feedback suppression system. (Modified from [5])

In Fig. 3.1, the input to the hearing-aid processing is $s(n)$, which is the sum of the desired input signal $x(n)$ and the feedback signal $f(n)$. The processed hearing-aid signal is $u(n)$. The amplifier impulse response is given by $a(n)$, the receiver impulse response by $r(n)$ and the microphone impulse response by $m(n)$. The signal in the ear canal is $y(n)$. The impulse response $b(n)$ includes both the acoustical (mainly contributed by a vent leading from the ear canal back to the microphone) and mechanical feedback, although the acoustical feedback usually dominates [11]. The acoustical feedback path through the vent tends to have a high-pass behaviour, and the un-amplified acoustical feed-forward path $c(n)$ through the vent from pinna to the ear canal tends to have a low-pass filter behaviour. It should be noted that the complete feedback path includes $m(n)$, $a(n)$, $r(n)$ and $b(n)$.

The transfer function of the hearing aid is as follows [12] [5]:

$$Y = \frac{X(C + HMAR)}{[1 - BC] - H[MARB]}, \tag{3.1}$$

where the capital symbol represents the z-transform of the corresponding quantity, for example $X = \mathcal{Z}\{x(n)\}$. Because C is a low-pass response and B is a high-pass response with reduced gain, one can safely assume that the product $|BC| << 1$. This assumption leads to a useful approximation of Eq.(3.1):

$$Y = \frac{X(C + HMAR)}{[1 - H[MARB]}. \tag{3.2}$$

As the denominator of Eq.(3.2) deviates from 1, the system transfer function deviates from the desired response. The danger occurs for a denominator close to zero at some frequency. If the denominator actually reaches zero, the transfer function for output Y involves a division by zero which indicates infinite output amplitude. A stable system is defined as one in which a finite input yields a finite output, therefore the hearing-aid system becomes unstable. In practice an infinite output level cannot be achieved given the finite power stored in the hearing-aid battery and saturations in the circuits, but the system tries to reinforce the output at the frequency where the gain is infinite, resulting in a continuous whistle output by the hearing aid [5]. A mathematical statement is that stability is ensured if the hearing-aid gain H is selected so that:

$$|HMARB| < 1, \tag{3.3}$$

at all angular frequencies. When $|HMARB| = 1$ and when the phase of $HMARB$ is an integer number of 2π, the system also oscillates.

3.2 Properties of the Feedback Path

Since the acoustical feedback usually dominates [11], in this work, only the acoustical feedback path is investigated. A typical acoustical feedback path represents a wave propagation path from the receiver to the microphone, which includes the effects of the hearing-aid receiver, earhook, PVC tubing, vent, leak, pinna, external acoustics, microphone, etc. These effects are described below.

3.2.1 Components of the Feedback Path

A typical magnitude frequency response of the omnidirectional microphone is fairly flat from 300 to 4000 Hz, but has a gradual roll-off below 300 Hz and a small resonance peak at around 5 kHz, and above 5 kHz the response falls smoothly [5] [8]. An example is shown in Fig. 3.2. By changing the size of the barometric relief hole in the diaphragm, the "corner frequency" at which the low-frequency response starts to decrease can be adjusted. The response peak is caused by the resonance of the diaphragm and the air within the microphone case and tube. One way to reduce the height of the resonance peak is to add damping to the resonance. Since the microphone response is usually adjusted to a relatively flat response, it is not the major contributor to the feedback path.

The response of the receiver is affected by the coupler and tubing used in the measurement. Fig. 3.3 shows the frequency responses of a hearing-aid receiver measured in the 2-cc, Zwislocki and 711 couplers and with different tubes. One important type of modifications of the receiver response to better match specific needs is to add damping to smooth the peaks in the frequency response either by placing a thin screen in the sound tube or through internal modifications. The effect of the receiver is an important contributor to the feedback path.

The vent configuration also has a large effect on the behaviour of the feedback path as shown in Fig. 3.4. The venting boosts the feedback path significantly. The study in [13] shows that when the vent size is changed from 2.1 mm to 3.1 mm, the feedback path is about 10 dB larger for frequencies above 1 kHz.

The earhook is used to retain the hearing aids over the ear and to conduct sound to tubing that is connected to the earhook [8]. The earhook is essentially a tube, therefore the effect of earhook is associated with resonances. To

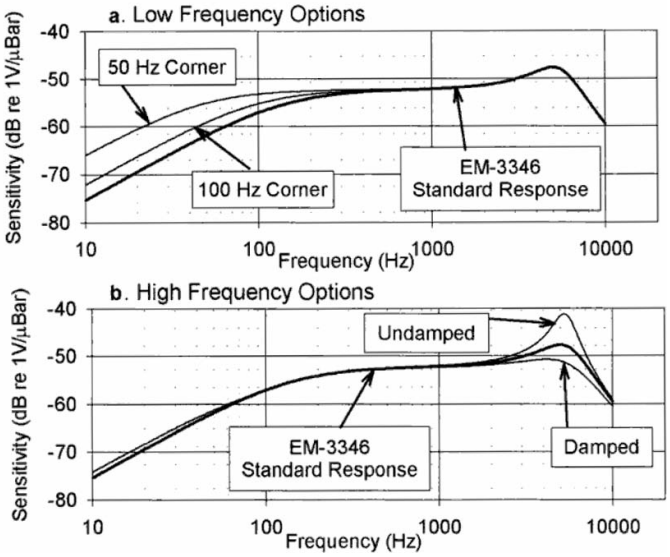


Figure 3.2: Sensitivity curves for a standard response microphone and several possible modifications of the standard response for the low (a) and high (b) frequency segment of the frequency response. (From [8])

reduce the sharp resonant peaks, different kinds of damping materials can be used in the earhook to achieve a smooth response.

The PVC tubing produces half-wave resonances. Different lengths and internal diameters result in different resonant peaks. A longer tube with a narrower internal diameter shifts the frequency peaks downward, and vice versa.

To our best knowledge, there is not much research on the nearfield effects of the head and pinna on the feedback path. However, the effects should not be underestimated especially at high frequencies when the wavelength is comparable to the size of the pinna. The effect is highly related to the location of the microphone. For different locations of the microphone, the difference in the magnitude response of the feedback path could be as large as tens of dBs at some frequencies.

The receiver and different tubing effects are the major contributors to the feedback path.

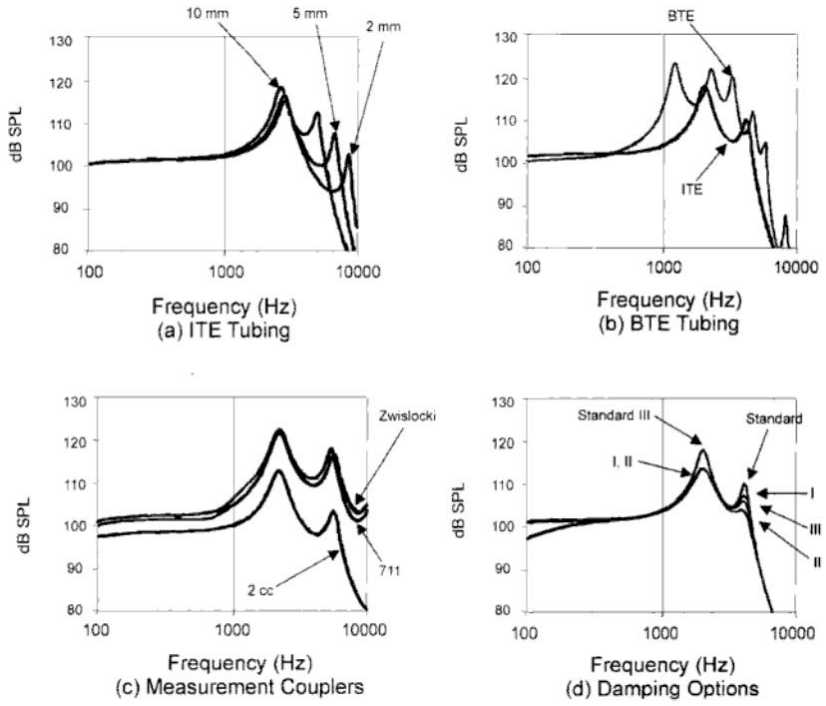


Figure 3.3: (a) Frequency response of a receiver with various lengths of coupling tubing attached. (b) Frequency response of a receiver with the longer tubing characteristic of a BTE aid. (c) Frequency response of a receiver measured in various couplers. (d) Frequency response of a receiver with different kinds and amounts of damping. Type I is a damping screen. Type II and III include internal damping. (From [8])

3.2.2 Properties of the Dynamic Feedback Path

The feedback path is subject to dynamic changes in real life. The causes of these changes fall into two categories: (a) changes in the hearing-aid fitting, including reinsertion of the hearing aid, jaw movement, smile, yawn, and cerumen; (b) changes in the external acoustics, including factors such as room reverberation, a hand or telephone handset near the ear. The variability of the feedback path under these conditions has been investigated intensively in previous studies [11] [13] [14].

To discuss the properties of the dynamic changes of the feedback path, a nominal feedback path is first defined. The nominal feedback path refers to the feedback path measured when the hearing aid is fitted properly to the subject who sits still without any external reflectors or enclosures near the hearing aid. The dynamic feedback path is analyzed by comparing the magnitude frequency response of the dynamic feedback path with that of the nominal

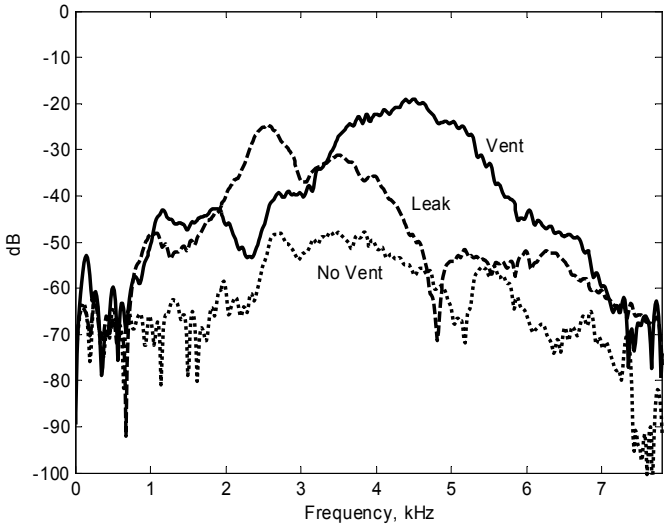


Figure 3.4: Feedback path magnitude frequency response measured on a dummy head for a BTE hearing aid. The earmold had no vent (dotted line), no vent but fitted loosely to ensure an acoustical leak (dashed line), or was vented (solid line). (From [5])

feedback path.

When the hearing aid and earmold are removed from the subject and reinserted to the ear in a correct way, the new nominal feedback path is found to be very similar to the first nominal feedback path in the study conducted by B. Rafaely *et al.* [11]. However, the study only investigated a specific type of hearing aids with earmold. In reality, the changes resulted from reinsertion for certain types of hearing aids could be larger if more freedom of moving is allowed during insertion.

When the hearing aid and the earmold are not fitted in a right way, for example, when the earmold is not fitted tightly to the ear, the measured feedback path is found to be different from the nominal feedback path as shown in the study [13] and [11]. The change is mainly at high frequencies. The mean change for frequencies between 2 and 5 kHz is 2.7 dB and 10.3 dB for BTE devices and ITC devices respectively. The maximum change is 5.7 dB and 21.4 dB for BTE and ITC devices respectively. Therefore, for ITC devices, a wrong placement can easily trigger the feedback whistling.

When the user opens his/her mouth during speaking, smiles or moves his jaw, the feedback path can also change. Rafaely *et al.* [11] found that the open mouth and smiling had little effect on the feedback path. However, it is also acknowledged that when face movements change the fitting of the hearing

aid and increase the leaks, they could have a more significant effect on the feedback path. The jaw movement also results in small average changes in the feedback path for both BTE and ITC devices, however, the standard deviation exceeds 6 dB for both ITC and BTE at some frequencies [13]. Therefore some users may have large problems with feedback during jaw movement.

Compared with the hearing-aid fitting, the changes in the external acoustics could bring much more significant changes in the feedback path. The typical changes in the external acoustics in real life include the following situations:

- when the user is wearing a knitted cap,
- when the user is hugging,
- and when the user is picking up a telephone handset.

A common observation in the studies [14] [11] [13] about these effects shows that the largest variation in the acoustical feedback path happens when a hand or telephone handset is placed near the ear. M. R. Stinson *et al.* investigated the effects of handset placed at different distances to the hearing aid. It is shown that for the BTE hearing aid, the feedback response is dramatically changed at high frequencies. From 100 mm distance to 0 mm, the changes reach 20 dB at 3 kHz. For ITE and ITC hearing aids, the changes are even more dramatic. For ITC hearing aids, the presence of the telephone handset increases the feedback path by nearly 20 dB at 2.5 kHz and the variation covers the whole frequency range. For ITE hearing aids, the general behaviour is similar to that of ITC hearing aids and shows a maximum change of 22 dB at around 3 kHz. The other two studies showed very similar results. Both [13] and [14] noticed the many peaks and valleys were present in the change of the magnitude frequency response and both attributed them to the reflection and refraction effects.

In this work, the telephone handset proximity effect was also investigated using a BTE hearing aid. The feedback paths were measured with a rigid surface sliding along the ear with a distance of around 3 cm. The main observation is that at high frequencies, the changes in the magnitude frequency response of the feedback path exhibit peaks and valleys, which agree with the previous studies mentioned above. The locations of the peaks and valleys justified the presence of reflections. For details, please refer to Paper I included in this dissertation.

The changes of the feedback path due to hug are similar to the telephone handset but the generally smaller [13].

The changes of the feedback path when the user is wearing a knitted cap exhibit a low-pass characteristic. A study by J. Hellgren *et al.* [13] showed that the knitted cap resulted in 3 dB higher magnitude response of the feedback path below 3 kHz on average for both BTE and ITC devices, and for frequencies above 4 kHz, the feedback response was 3 dB lower with the BTE devices and relatively unchanged with ITC devices.

3.3 Modelling of the Feedback Path

As mentioned in Sec. 3.2, the feedback path is composed of many components of different characteristics. In this work, it is therefore suggested to divide these components into three categories:

- Category I: Device type dependent components. For a specific device, the effects of the components in this category are invariant or only slowly varying, and are independent of the users and the external acoustical environment. These components include the hearing-aid receiver, microphone, tube attached to the receiver inside the hearing-aid shell, etc.
- Category II: User dependent components, which include the PVC tubing, earmold, pinna, etc. The change of the hearing-aid fitting is caused by the change of the components in this category. The change is usually slow but could be fast, for example, when the user moves his/her jaw quickly.
- Category III: External acoustical environment dependent components. The change of the components in this category can be very rapid and dramatic, for example, when the user picks up a telephone handset.

The components in Category II and III cause a large inter-subject variability in the feedback path and a large variation of the feedback path over time.

The modelling of the feedback path is generally studied from two perspectives. Some studies have focused on modelling each component of the feedback path theoretically by using the equivalent electro-acoustical models [15] [16] [17] or numerically by using methods such as boundary element calculations [14]. However, it is very difficult for these models to be precise due to oversimplification of the problem or unknown/imprecise parameters. In

addition, they only model the components in Category I and some of the components in Category II.

The other studies have focused on modelling the overall feedback path. This kind of models has been proposed for the purpose of adaptive feedback cancellation. It can model the nominal feedback path as well as the changes in the feedback path on-line due to its simplicity and accuracy. In general, two models have been proposed. In the first model, the overall feedback path is represented with an FIR filter [18] [19], which can be adaptive to track the changes in the feedback path. This model is referred to as "direct model". The second model, which is referred to as "initialization model", incorporates a fixed filter (usually an IIR filter) and an adaptive FIR filter [20] [21]. The fixed model represents the slowly varying portion of the feedback path (microphone, receiver, etc.), whereas the adaptive filter represents the rapidly varying portion. In practice, the way to obtain the coefficients of the fixed filter is to measure the feedback path using a probe signal for each individual user when the hearing aid is fitted, and fit the fixed filter to the measured response of the feedback path. Compared with the first model, the second model yields a shorter adaptive FIR filter, a faster converge speed and a smaller computational load. However, it requires an additional procedure in the hearing-aid fitting and the fixed filter actually includes some of the fast varying components in Category II and III.

In this work, the modelling of the feedback path is investigated from the second perspective. An ideal feedback path model is suggested, which consists of a fixed model representing the invariant components (mainly in Category I), a slowly varying model and a fast varying model. The slowly varying model is used to model the slow changes in the components in Category I (due to aging or drifting), II (due to the slow changes in the hearing-aid fitting) and III (due to the slow changes in the acoustical environment). The fast varying model is used mainly for modelling the rapid and dramatic changes in the external acoustics, for example, when the user picks up a telephone handset.

For the fixed model, the extraction of the "invariant part" of the feedback path is investigated, which is independent of the users and the acoustical environment. For different types of devices, the "invariant part" is obviously different. For the same type of devices, the "invariant part" could also be different from device to device in reality due to the variation within the batch of the components. The investigation mainly focuses on how to extract the invariant model for a given type of hearing aids. The idea is to measure a number of feedback paths using the same type of hearing aids on different

users. The invariant part of the feedback path can then be regarded as the common part of these measured feedback paths. Two extraction methods, i.e. a Common-Acoustical-Pole Zero model-based approach and an Iterative-Least-Square-Search based method, are used and a combination of the two methods is proposed. The results show that the methods can extract the invariant part effectively and the proposed method gives the best overall performance. The study also shows that the invariant part of the feedback path is not trivial in practice although variations within the batch of components exist. Compared with the fitting procedure used in the "initialization model", these extraction methods can be used off-line in the production phase for a specific type of hearing aids to avoid the additional individual fitting procedure. In addition, it only identifies the invariant part instead of the overall nominal feedback path. For details, please refer to Paper III included in this dissertation.

One type of the fast varying models is also proposed in this work. The capabilities of the "direct model" and the "initialization model" for modelling the rapid feedback changes with a reflecting object placed close to the ear, which is regarded as one of the situations resulting in the most significant change in the feedback path, are first investigated. A novel reflection model is proposed based on the observations that these difficult situations usually come along with reflections. The reflection gains and delays are estimated using a weighted Fourier transformation and relaxation (WRELAX) based method. The results using the measured feedback paths of a commercial hearing aid show that the reflection model has a better ability of capturing the dynamic feedback path in this difficult real-life situation and is superior to the existing two models in terms of MSE and MSG. To achieve the same level of MSE or MSG, the number of components required in the proposed model is fewer than the orders of the FIR filter required in the two existing models. The study also shows the minimum order of the adaptive filter in the two existing models to achieve a certain MSG in the difficult dynamic situations, which could serve as a useful indication in practice for choosing the order of the adaptive filter. It is also found that for the reflection model, eleven reflection replicas are enough to model the dominant characteristics of the proximity effect on the dynamic feedback paths. For details, please refer to Paper I included in this dissertation.

CHAPTER

FOUR

Feedback Suppression Techniques

To address the feedback problem, many approaches have been proposed in the last two decades. The existing techniques can be classified into forward-path suppression and feedback cancellation [1]. In this Chapter, these approaches will be reviewed. The new methods proposed in this work will be described briefly. In the end, the way to use the new feedback path models proposed in Chap. 3 for feedback suppression will be discussed.

4.1 Forward-path Suppression

One way to prevent the system from whistling is to modify the forward-path H of the hearing aids in such a way that it is stable in conjunction with the feedback path B [1]. This approach is called "forward-path suppression".

Among these approaches, the simplest one to reduce the occurrence of the feedback problems is to reduce the gain in order to satisfy Eq.(3.3). The gain does not need to be reduced everywhere. It only needs to be reduced in the critical frequency region where the feedback problems are expected to occur [5]. There are two ways to reduce the gain. One way is to insert the notch filter into the forward-path processing according to the peaks of the system response identified in advance. The other way is to reduce the gain when it is necessary and apply the gain reduction at the frequency regions where system stability occurs. A detector is usually used to identify the situations where the gain reduction should take effect. This detector can be based on the spectral of the input signal [22] or other features of the system/signals. The

major drawback of gain reduction is that it compromises the amplification of the hearing aids and therefore could reduce the speech intelligibility [23].

Other approaches include using delay or phase modulation and frequency shifting in the forward-path [24] [25] [26].

The increase in MSG with forward-path suppression techniques has generally been found limited [1]. In addition, the forward-path suppression techniques either compromise the sound quality or limit the amplification desired by the users.

4.2 Adaptive Feedback Cancellation

A more promising way of feedback suppression is feedback cancellation as illustrated in Fig. 4.1.

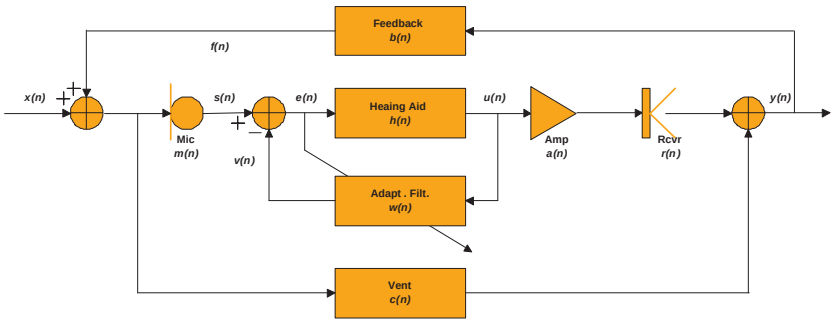


Figure 4.1: Block diagram of a generic feedback cancellation system. (Modified from [5]) The feedback path is modelled by an adaptive filter $w(n)$, and $v(n)$ is the estimate of $f(n)$. It should be noted that the complete feedback path includes $m(n)$, $a(n)$, $r(n)$ and $b(n)$.

The idea is to use an adaptive filter to model the feedback path and subtract the estimated feedback signal from the signal input to the hearing aids.

The transfer function of the hearing aid can be obtained by considering the feedback path model in Eq.(3.1) [23]:

$$Y = X\frac{C + H[MAR + WC]}{[1 - BC] - H\{MARB - W[1 - BC]\}}. \tag{4.1}$$

If it is assumed that $|BC| << 1$, Eq.(4.1) is simplified to:

$$Y = X\frac{C + H[MAR + WC]}{1 - H[MARB - W]}. \tag{4.2}$$

The stability is guaranteed if

$$|H[MARB - W]| < 1. \tag{4.3}$$

If W is a good model of the actual feedback path $MARB$, the MSG of H can be made larger. If there is a perfect match between modelled and real feedback path, the feedback signal $f(n)$ will be cancelled completely and the system will be stable for any amount of amplification. On the contrary, if there is much difference between them, the MSG will be reduced significantly.

Different adaptation methods have been proposed for updating the feedback path model $w(n)$. They can be divided into two classes: algorithms with continuous adaptation and algorithms with a non-continuous adaptation [1]. In the second class of algorithms, the feedback path model is adapted only from time to time according to the stability of the system or the characteristics of the input signal. For the former case, the problem is that it does not activate the adaptation until instability has occurred. For the latter case, the adaptation may halt waiting for a proper signal while the feedback path is rapidly changing. This will result in instability easily.

In the adaptive feedback cancellation, the adaptive part of the feedback path model is usually an FIR filter because the adaptive IIR filter suffers from the problem of computational complexity, instability, and local minima [22]. The continuous adaptation based feedback canceller can use any standard on-line adaptation methods to update the adaptive filter, such as the LMS algorithm, the NLMS algorithm, and the RLS. However, due to the requirement of low complexity for algorithms in hearing aids, simple adaptation algorithms are usually used. The simplest form of adaptation is the LMS algorithm [27], in which the following cost function is minimized:

$$J(n) = |e(n)|^2. \tag{4.4}$$

The adaptive FIR filter coefficients are updated as below:

$$\begin{aligned} \hat{\mathbf{w}}(n+1) &= \hat{\mathbf{w}}(n) + \mu(n)\mathbf{u}(n)e(n), && (4.5)\\ \hat{\mathbf{w}}(n) &= [\hat{w}_0(n), \hat{w}_1(n), \ldots, \hat{w}_{p-1}(n)]^T, && (4.6)\\ \mathbf{u}(n) &= [u(n), u(n-1), \ldots, u(n-p+1)]^T, && (4.7) \end{aligned}$$

where $\mu(n)$ is the step size which could be time-varying, $\hat{\mathbf{w}}(n)$ is the vector form of the coefficients of the adaptive FIR filter of order p, and $e(n)$ is the error signal as shown in Fig. 4.1.

With a wide sense stationary input and a proper step size, the adaptive filter will converge towards the Wiener solution. Ideally, the adaptive filter coefficients converge to the real feedback path response so that the error signal

$e(n)$ equals the desired input signal $x(n)$. However, the Wiener solution is not necessarily close to the real feedback path response even when the order of the filter is sufficiently high and quantization error of the filter coefficients is negligible. The deviation of the Wiener solution from the true feedback path is the so-called "bias problem", which is explained in the next section.

4.3 Bias of the Adaptive Feedback Cancellation

Due to the correlation between the input and output of the hearing aid, the Wiener solution does not actually approximate the true feedback path. The Wiener solution to the minimization of the mean square of the error signal $e(n)$ is:

$$\begin{aligned}
\hat{\mathbf{w}} &= \mathbf{R}_{\mathbf{uu}}^{-1}\mathbf{r}_{\mathbf{u}s} && (4.8)\\
&= \mathbf{R}_{\mathbf{uu}}^{-1}\mathbf{r}_{\mathbf{u}f} + \mathbf{R}_{\mathbf{uu}}^{-1}\mathbf{r}_{\mathbf{u}x}, && (4.9)\\
&= \mathbf{R}_{\mathbf{uu}}^{-1}\mathbf{R}_{\mathbf{uu}}\mathbf{b} + \mathbf{R}_{\mathbf{uu}}^{-1}\mathbf{r}_{\mathbf{u}x} && (4.10)\\
&= \mathbf{b} + \mathbf{R}_{\mathbf{uu}}^{-1}\mathbf{r}_{\mathbf{u}x}, && (4.11)\\
\mathbf{R}_{\mathbf{uu}} &= E\left\{\mathbf{u}(n)\mathbf{u}^T(n)\right\}, && (4.12)\\
\mathbf{r}_{\mathbf{u}s} &= E\left\{\mathbf{u}(n)s(n)\right\}, && (4.13)\\
\mathbf{b} &= \left[b(0), b(1), \ldots, b(p-1)\right]^T, && (4.14)
\end{aligned}$$

where $\mathbf{r}_{\mathbf{u}x}$ and $\mathbf{r}_{\mathbf{u}f}$ are defined similarly as in (4.13). It should be noted that the microphone, receiver and amplifier responses $m(n)$, $r(n)$ and $a(n)$ are dropped from now on for simplicity and are all absorbed in the feedback path $b(n)$ of the length p (the same as the length of the adaptive filter $w(n)$).

In Eq.(4.11), the term $\mathbf{R}_{\mathbf{uu}}^{-1}\mathbf{r}_{\mathbf{u}x}$ represents the bias of the estimate, which is related to the correlation between the desired input signal $x(n)$ and the processed hearing-aid signal $u(n)$. The magnitude of the bias depends strongly on the decaying speed of the autocorrelation function of $x(n)$, the forward-path delay and nonlinearity in the hearing-aid process H. When $x(n)$ is a white noise sequence, the estimate is not biased. When $x(n)$ is tonal, the bias problem is particularly serious because the correlation function does not drop off.

The introduced bias severely limits the performance of the feedback canceller especially when tonal signals or musical signals are input to the hearing aids. In these situations, the magnitude of the adaptive FIR filter coefficients are abnormally large and drifted far from the true feedback path. The excessively large filter coefficients, when combined with the hearing-aid amplification in

the forward-path processing H, can result in system instability or colouration of the output signal due to large undesired changes in the system frequency response [23].

4.4 Bias Reduction Techniques

Bias reduction has been one of the main topics of adaptive feedback cancellation. During the past two decades, various approaches have been proposed to decorrelate the input and output of a hearing aid to reduce the bias in the estimate of the feedback path.

One well-known decorrelation approach introduces a delay in the hearing-aid processing (or the feedback cancellation path) to reduce the correlation between $x(n)$ and $u(n)$. It has been shown in [28] that for a coloured noise input, the insertion of delay in the hearing-aid processing significantly improves the accuracy of feedback path modelling, while the insertion of a delay in the feedback cancellation path provides smaller benefit. However, the delay introduced in the hearing aids should be kept small to avoid disturbing artifacts such as comb filtering [29]. Moreover, for tonal signals, a delay will not help much to reduce the correlation.

Another kind of decorrelation approaches uses nonlinearities in the forward-path processing. These methods include frequency shifting [30], time-varying all-pass filter [31], etc. Since all the nonlinear methods degrade sound quality to some extent, a trade-off between the performance of feedback cancellation and sound quality is usually involved.

Alternatively, a probe signal, usually a noise signal, can be added to the receiver input $u(n)$ [32]. To maintain sound quality, the probe signal should be inaudible, and its level therefore has to be much lower than that of the original receiver input signal. The bias reduction achieved with such a weak probe signal is very small.

The bias of the adaptive feedback cancellation can also be reduced by controlling the adaptation. One way is to change the speed of adaptation. For example, when a sudden change in the feedback path is detected or when the input signal is white noise, the adaptation speed is increased, and otherwise, the adaptation speed is kept small to reduce the distortion of the desired signal. Another way is to constrain the adaptation so that the filter coefficients do not drift too far away from the assumed "correct" coefficients (also called "reference filter") [23]. This technique needs an additional initialization dur-

ing the fitting of hearing aids to provide a reasonably good reference for the feedback cancellation since in practice there is no "correct solution" available. In the initialization a noise-like periodic excitation sequence is used to measure the feedback path response of the hearing aid inserted in the user's ear.

Recently proposed approaches use the filtered-X adaptation, in which the two signals $u(n)$ and $e(n)$ pass through the same filter before they are used to update the adaptive filter $w(n)$ (cf. Eq.(4.5)). In [33], the two signals pass through the same band-pass or high-pass filter to limit the feedback cancellation signal in the frequency band where possible oscillation may occur. In [20], the two signals pass through the same adaptive filter, which removes the low-frequency tonal components in the error signal $e(n)$. In this way, the sensitivity to low-frequency tonal inputs of the feedback canceller is reduced. A similar approach is proposed in [19] from a system identification point of view, in which the two signals $u(n)$ and $e(n)$ are also filtered by the same adaptive filter before being used to update the feedback path model. The coefficients of the adaptive filter are obtained from linear prediction of the error signal $e(n)$ to remove the correlation between consecutive samples of the signal and hence reduce the autocorrelation function. It has been proved in [19] that when the desired input signal $x(n)$ is an AR random process and the order of the feedback path model is sufficient, the bias is removed completely.

The most effective feedback cancellation is probably a combination of the aforementioned techniques. An example of such a feedback canceller is illustrated in Fig. 4.2, which is referred to as "modern feedback canceller".

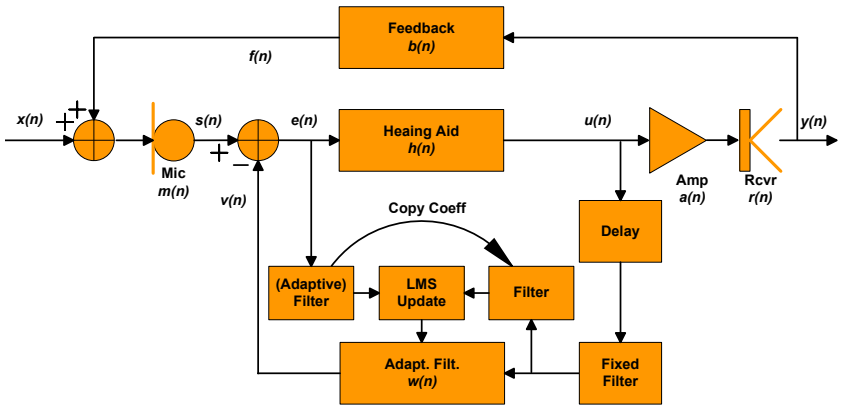


Figure 4.2: Block diagram of a modern feedback cancellation system. (Modified from [5]) It should be noted that the complete feedback path includes $m(n)$, $a(n)$, $r(n)$ and $b(n)$.

The "delay" block is introduced to match the physical delay in the wave propagation from the receiver to the microphone. The delay and the fixed filter represent aspects of the physical feedback path that are expected to change very slowly over time [5]. The coefficients of the fixed filter is obtained by an individual fitting procedure as mentioned above in the "initialization model". The adaptive filter $w(n)$ models the rapid-varying portion, which can be constrained by the reference filter [23]. Two identical filters are used to modify the inputs to the LMS update (filtered-X adaptation), which can be fixed filters [21] [33] [12] or adaptive filters using linear prediction [19].

This diagram can be regarded as one type of generic modern feedback cancellers in the digital hearing aids since most of the feedback cancellers based on the state-of-the-art techniques (filtered-X adaptation, constrained adaptation, linear prediction, etc.) are special variants of the structure in Fig. 4.2.

In this work, two new approaches are proposed to reduce the bias in the feedback cancellation. Both methods combine the traditional approaches and new approaches proposed recently.

The first approach uses injected noise to reduce the bias in the prediction error method based adaptive feedback canceller (PEM-AFC) proposed in [19], which is essentially a filtered-X adaptation based feedback canceller. Although it is proved in [19] that the PEM-AFC can eliminate the bias when certain conditions are met, the bias still remains in many situations, for example when the desired input signal is voiced speech. Therefore, a reduction of the remaining bias is necessary. The study in [19] showed that noise injection could improve the performance of the PEM-AFC further. However, the injected noise used in that work was an audible speech-shaped noise and therefore compromised the sound quality. Two nearly inaudible noises are proposed to use with PEM-AFC in our study: a masking noise, which is tailored to the hearing-aid application, and a linear prediction based noise, which is especially efficient for feedback cancellation with linear prediction. Simulation results show that noise injection approach is more effective in the PEM-AFC than in the traditional feedback cancellation without using the filtered-X adaptation (shown in Fig. 4.1), and the injection of these two kinds of noises can further reduce the feedback estimation error of the PEM-AFC by 1-4 dB and/or increase the stable gain by 3-4 dB, depending on the characteristics of the input signal. For details, please refer to Paper IV included in this dissertation.

The second approach uses a band-limited linear predictive coding (BLPC) vocoder to reduce the bias. The idea is to replace the hearing-aid output with a synthesized signal, which sounds perceptually similar to or possibly even

the same as the original signal but is statistically uncorrelated with the desired input signal. To obtain effective decorrelation in the BLPC vocoder, impulse trains are not used for excitation as in conventional LPC-based vocoders during voiced speech. Instead, a white noise sequence is used to drive the estimated signal model to generate the synthesized signal. Based on the fact that the magnitude of the frequency response of the feedback path is usually much higher in the high-frequency region and that the adaptive feedback cancellation usually breaks down at high frequencies, the signal replacement is performed at high frequencies to improve the performance of the adaptive feedback canceller in the critical frequency region and also to reduce the degradation in sound quality. According to our subjective evaluation, the sound quality is very well preserved for speech. For many musical signals with only a few peaks sparsely spaced at the high-frequency spectrogram, the sound quality is not degraded very much either. For musical signals with many peaks in the spectrogram, the degradation of sound quality is expected.

The proposed BLPC vocoder can be used with a conventional AFC to yield a BLPC-AFC, which reduces the long-term bias. Moreover, the BLPC-AFC can be further combined with the filtered-X adaptation to get a BLPC-FxAFC, which can effectively reduce the short-term bias. The proposed BLPC-FxAFC can also be regarded as a modified version of the PEM-AFC combined with the BLPC vocoder. The simulation results show that the use of the proposed BLPC vocoder can effectively reduce the bias and the misalignment between the real and the estimated feedback path. The BLPC-FxAFC method has the best performance for all the test signals. For details, please refer to Paper II included in this dissertation.

It should be noted that in this work, all the adaptive feedback cancellers operate in the time domain for simplicity. It is always possible to translate them into frequency-domain implementations.

4.5 Feedback Path Models used in the Adaptive Feedback Cancellation

The adaptive feedback canceller implicitly assumes a certain model of the feedback path. In Sec. 3.3, these models have been discussed, and two new feedback path models are proposed. The the feedback path models used in the adaptive feedback cancellation is discussed in this section.

As mentioned in Sec. 3.3, it is suggested in this work that the ideal feedback

path model, based on the characteristics of the different components of the feedback path, should consist of three parts: a fixed model, a slowly varying model and a fast varying model. The model used in the modern feedback canceller in Fig. 4.2 can be regarded as a simplified version of this ideal model. The differences between the ideal model and the model used in the modern feedback canceller are:

- In the model used in the modern feedback canceller, the fixed filter in practice requires an individualized fitting procedure and models not only the invariant portion but also some variant characteristics.
- In the model used in the modern feedback canceller, the slowly varying model and the fast varying model are not separated.

To address the first problem, methods to extract the invariant part from the feedback paths are proposed. This extraction can be done off-line for each type of hearing instruments without any individual fitting procedure and the model extracted can be used in the fixed filter directly.

The second problem is more complicated. The drawback of mixing the slowly and fast varying models in the adaptive feedback cancellation is that the adaptation speed is hard to control. To minimize the distortion during normal daily use, the adaptation speed should be small. On the other hand, to track the fast changes, for example, due to the rapid changes in the external acoustics, the adaptation should be accelerated. One way to address this dilemma is to use the ideal model and divide the adaptive filter into two cascaded adaptive filters with different adaptation speeds. However, the feasible scheme for this separation has not been investigated in this work due to the limited time. Some alternative ways have been proposed in the previous studies. For example, a fast adaptive filter and a slow adaptive filter can be used in parallel, and a switch controls which one should be active as used in [34]. The adaptation speed can also be altered according to the conditions of the system, for example according to the prediction error power normalized by the input power in the PEM-AFC-like feedback cancellers [35]. In the feedback canceller with adaptation constrained by the reference filter, the reference filter can slowly adapt to absorb the slow changes in the feedback path, whereas the adaptive feedback path model itself may still adapt fast but does not deviate too much from the adaptive reference filter [36].

The reflection model proposed in this work can be regarded as one type of the fast varying models. However, due to the limited time for this work, the on-line update scheme for direct use of this model in the adaptive feedback

cancellation has not been investigated.

CHAPTER

FIVE

Conclusions and Suggestions for Further Research

In this dissertation, the feedback whistling problem with digital hearing instruments has been investigated. The work mainly addresses the feedback problem from two perspectives:

- Feedback path modelling
- Feedback suppression techniques

5.1 Conclusions

The properties and the modelling of the feedback path are first discussed. Along the propagation path of the feedback signal, the components of the feedback path are categorized into three groups: device dependent components, user dependent components and external acoustical environment dependent components. According to the characteristics of different components, an ideal feedback path model is proposed, which consists of a fixed model, a slowly varying model and a fast varying model. Methods for extracting the fixed model are proposed and proved to be effective. Based on the observations that rapid changes in the feedback path usually come along with reflections, a fast varying model based on reflection assumptions is proposed. The results show that the new model is very efficient in representing the dynamic feedback path. The extracted fixed model can be used directly in the adaptive feedback cancellation. The reflection model still needs a new adaptation scheme to be used directly in the feedback cancellation.

This dissertation also investigates the feedback suppression techniques, which can be divided into two classes: forward-path suppression and feedback cancellation. The focus of this work is the algorithms in the second class, which is proved to be more promising. One of the biggest issues with the adaptive feedback cancellation is the so-called "bias problem". To reduce bias, two approaches are proposed, both of which are based on the feedback canceller with filtered-X adaptation. The first approach injects two kinds of nearly inaudible noises to help reduce the bias. It is shown that the noise injection in the filtered-X adaptation based feedback cancellers is more effective than in the traditional feedback cancellers. The second approach uses a linear predictive coding based vocoder to synthesize the output signal of the hearing aid so that the synthetic signal sounds similar to or even the same as the original signal but is statistically uncorrelated with the desired input signal. The synthesis is bandlimited to reduce the effects on sound quality. Simulation results show that the proposed vocoder can effectively reduce the bias and the misalignment between the real and the estimated feedback path. When combined with the filtered-X adaptation based feedback canceller, this vocoder reduces the misalignment even further.

5.2 Suggestions for Future Research

In Sec. 3.3, a reflection model for the dynamic feedback path is proposed. However, to use it for the feedback cancellation, an adaptation scheme should be investigated. Since the delays are fractional, it will be easier to adapt in the frequency domain. The finding of optimal delays is a nonlinear problem and therefore an approximate solution has to be found.

In the band-limited linear predictive coding based feedback canceller, the sound quality and speech intelligibility have only been assessed by the authors. A clinical trial and/or objective measure is still needed in the future for this approach.

Compared with the suggested ideal feedback path model, the model used in the existing feedback cancellers mixes the slowly varying and fast varying models. Therefore, the methods for splitting these two models should be investigated.

The biggest challenges for feedback suppression still remain:

- The fast track of the feedback path under adverse conditions, for example, when the user brings the telephone handset to his/her ear.

- The bias problem in the adaptive feedback cancellation when the input is a narrow-band signal.

These problems are still the focuses of the future research on the feedback suppression in the digital hearing aids.

Paper I

Using a Reflection Model for Modeling the Dynamic Feedback Path of Digital Hearing Aids

G. Ma, F. Gran, F. Jacobsen and F. T. Agerkvist

Journal of the Acoustical Society of America, Vol. 127, No. 3, pp. 1458-1468, March 2010

Using a reflection model for modeling the dynamic feedback path of digital hearing aids[a)]

Guilin Ma[b)] and Fredrik Gran
Research Group, GN ReSound A/S, Lautrupbjerg 9, 2750 Ballerup, Denmark

Finn Jacobsen and Finn Agerkvist
Department of Electrical Engineering, Acoustic Technology, Technical University of Denmark, Building 352, 2800 Kgs. Lyngby, Denmark

(Received 6 May 2009; revised 16 September 2009; accepted 13 December 2009)

Feedback whistling is one of the severe problems with hearing aids, especially in dynamic situations when the users hug, pick up a telephone, etc. This paper investigates the properties of the dynamic feedback paths of digital hearing aids and proposes a model based on a reflection assumption. The model is compared with two existing models: a direct model and an initialization model, using the measured dynamic feedback paths. The comparison shows that the proposed approach is able to model the dynamic feedback paths more efficiently and accurately in terms of mean-square error and maximum stable gain. The method is also extended to dual-microphone hearing aids to assess the possibility of relating the two dynamic feedback paths through the reflection model. However, it is found that in a complicated acoustic environment, the relation between the two feedback paths can be very intricate and difficult to exploit to yield better modeling of the dynamic feedback paths. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3290989]

PACS number(s): 43.66.Ts, 43.20.El, 43.60.Ac, 43.60.Uv [AJZ] Pages: 1458–1468

I. INTRODUCTION

Feedback is a severe problem in hearing aids, which limits the maximum gain that can be achieved.[1] The acoustic feedback path is the most significant contributor to the feedback signal although electrical and mechanical paths also exist.[2] A widely adopted approach to acoustic feedback suppression is feedback cancellation, where the acoustic feedback signal is estimated and then subtracted from the input signal to remove feedback.[3] The maximum stable gain (MSG) that can be obtained by feedback cancellation depends on how accurately the feedback path is estimated. A perfect match between the modeled and the real feedback path will cancel the feedback signal completely, and the system will be stable for any amount of amplification.[4]

A typical acoustic feedback path in the hearing aids includes the effects of the hearing-aid amplifier, receiver, microphone, the acoustics of the vent or leak, as well as the external acoustics. Except for the fitting and external acoustics, all the parts are nearly constant or very slowly varying with a time frame of weeks or years[5] when the user has put on the hearing aid. Assume that the user stays still in a stationary acoustic environment, and the feedback path will be almost static. This feedback path is referred to as a "static feedback path." However, in practice, the feedback path is usually subject to dynamic changes. The causes of these changes fall into two categories: (a) changes in the hearing-aid fitting, including jaw movement, smile, yawn, and cerumen; (b) changes in the external acoustics, including the factors such as room reverberation, a hand, or telephone handset near the ear. The variant feedback path due to these changes is called "dynamic feedback path." In this paper, only the dynamic feedback path for a specific position along the changing course is studied. The temporal characteristics, such as the changing rate at that position, are not addressed.

The properties of feedback paths have been investigated intensively in previous studies,[5,2,6] where the effect of vent size, the variability of feedback due to smiling, and handset proximity, were discussed. A common observation in these studies is that the largest variation in the acoustic feedback path was found when a hand or telephone handset was placed near the ear. For behind-the-ear (BTE) devices, the maximum variation can go up to 20 dB, whereas for in-the-ear (ITE) device, it can even go up to 27 dB in the 2–5 kHz frequency range compared with the static feedback path.[6]

The modeling of the feedback path has also been discussed intensively. However, the majority of the previous work has focused on the static feedback path, trying to model each static component, such as cavities and transducers, by an equivalent electroacoustic model.[7–9] For dynamic feedback path modeling, there is not very much research up to the present. One study that is relevant measured the variability of feedback path under various conditions and introduced an uncertainty bound model to calculate the robust stability condition for the hearing-aid system.[2] It provided knowledge about one aspect of the dynamic feedback path properties, i.e., the extent of variability, but did not discuss other acoustic properties and did not investigate the modeling issue either. Another relevant study dealt with feedback cancellation

[a)]Portions of this work were presented at the 34th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) in April 2009.

[b)]Also at Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark. Author to whom correspondence should be addressed. Electronic mail: gm@elektro.dtu.dk

 0001-4966/2010/127(3)/1458/11/$25.00

in room reverberation.[10] The focus of that study was the performance of feedback cancellation instead of the properties of the feedback path and its modeling in reverberation.

To track the dynamic aspect of feedback paths in the real world, feedback cancellation in many hearing aids today includes an adaptive feedback path model, where a simple adaptive finite-impulse-response (FIR) filter is used instead of an infinite-impulse-response (IIR) filter, since adaptive IIR filter suffers from the problem of computational complexity, instability, and local minima.[11] Thus, the estimate of dynamic feedback path is essentially a procedure of updating the coefficients of the adaptive FIR filter.

In general, two models with different ways of using the adaptive FIR filter have been proposed to model the dynamic feedback paths. In the first model, such as the systems proposed in some previous studies,[12,13] the overall feedback path is represented with the FIR filter. This model is referred to as the "direct model." The second model, such as the systems proposed in other studies,[3,14] incorporates a fixed model and an adaptive FIR filter. The fixed model represents the slowly varying portion of the feedback path (microphone, amplifier, and receiver), whereas the adaptive filter represents the rapidly varying portion (mainly the change of fitting and external acoustics). The fixed model can be either captured by an initialization,[3] referred to as an "initialization model" in this paper, or roughly approximated as a high-pass filter.[14] Compared with the direct model, the initialization model generally needs a shorter FIR filter and is computationally more efficient but requires an additional procedure in the hearing-aid fitting.

No matter which model is used, effective feedback cancellation requires that the FIR filter should adapt fast enough given the variability of the feedback paths and should be long enough to model all of the salient features.[10] The feedback path varies at different rates under different conditions. For example, jaw movements will produce changes with a time frame of seconds/minutes.[15] The picking up of a telephone handset changes the feedback path with a time frame of milliseconds/seconds. Production and disposal of cerumen will result in changes with a time frame of weeks and months, while aging causes changes with a time frame of years.[5] Usually the adaptation speed is not a problem even in the extreme cases as long as a proper adaptation scheme, for example, frequency-domain adaptation, is used. However, in practice, to maintain sound quality and minimize the artifacts resulted from adaptation errors when a spectrally colored signal is inputted,[13] the adaptation usually has to slow down and is not fast enough in some occasions. These adaptation problems, although prominent and complicated in the feedback cancellation system, are more related to the on-line adaptation scheme and the characteristics of the input signals, instead of the models of the feedback paths. The focus of this paper is on the modeling. Therefore, these adaptation issues are not addressed in this paper.

As for the sufficient length of the FIR filter in the model, the natural questions are as follows: how many orders of the FIR filter would be "enough" in various dynamic situations, and is there any other model that is more effective in modeling the dynamic feedback paths? In order to address these issues, the characteristics of the dynamic feedback path in addition to variability should be examined. This paper investigates these properties for the most adverse proximity situation that happens frequently and challenges the feedback cancellation most. This paper also describes a novel reflection based model (proposed in a previous publication[16]) in detail and extends the previous work thoroughly. The basic idea of the reflection model is to assume that the dynamic feedback path consists of multiple propagation paths and estimate the delays and gains associated with these paths directly. It differs from the two existing models in that the reflection model has an underlying physical assumption, and is more generalized by allowing fractional delays in the model, which will be described in more detail in Sec. II. The effectiveness of the new model is compared with the existing two models mentioned above, i.e., the direct model and the initialization model, using data from measurements. The comparison shows that the proposed model is superior in terms of mean-square error (MSE) and MSG.

For a hearing aid with dual microphones, due to the short distance between the two microphones (normally less than 2 cm), the two feedback paths are similar in most situations. This paper investigates the possibility of relating the two dynamic feedback paths using the initialization model and the reflection model. It is shown that in a complicated acoustic environment, the relation between the two feedback paths can be very intricate and difficult to exploit to yield better modeling of the dynamic feedback paths.

The outline of this paper is as follows: in Sec. II, two traditional models are explained and a new reflection model is proposed. In Sec. III, the measurement configuration and procedure are described. The properties of the dynamic feedback paths are discussed in Sec. IV. In Sec. V, the performance of the proposed model is compared with the existing two models and the possibility of using this model for relating the two dynamic feedback paths of a dual-microphone hearing aid is explored based on the measured data. Concluding remarks and directions for future work are given in Sec. VI.

II. MODELS FOR THE DYNAMIC FEEDBACK PATH

The general diagram of adaptive feedback cancellation is depicted in Fig. 1. It should be noted that the impulse responses of the microphone and receiver have been included in the impulse response of the feedback path $b(n)$. Although there are various feedback models as mentioned in Sec. I, their estimation of $b(n)$ can all be represented in a compact form $\hat{b}(n,\theta)$, where n denotes the discrete-time index and θ is a particular parameter set of a model. For example, when the feedback path is modeled by a FIR filter, θ represents the FIR coefficients. The principle of feedback cancellation is to adjust the parameters θ in the feedback model so that the modeled feedback path $\hat{b}(n,\theta)$ approximates the true feedback path $b(n)$. The output $v(n)$ is the instantaneous estimation of the feedback signal $f(n)$ and is subtracted from the input signal $s(n)$ to remove the feedback.

In principle, the impulse response $b(n)$ has infinite duration. However, the amplitude of $b(n)$ decays very fast, as

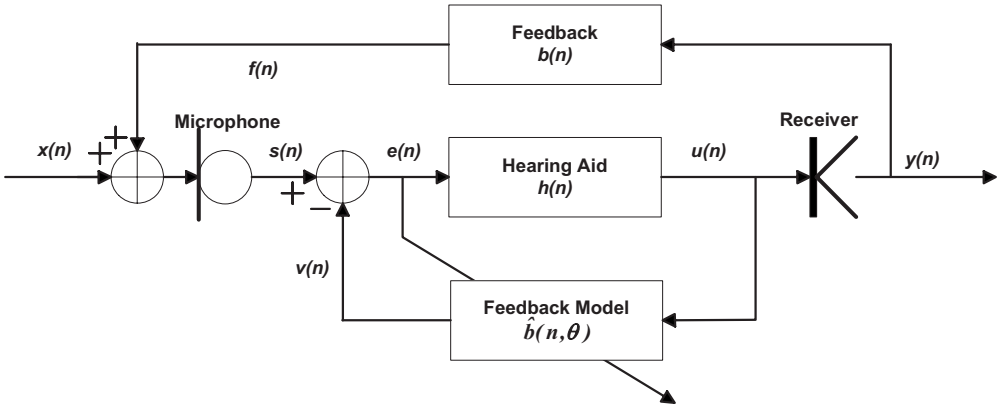


FIG. 1. General diagram of adaptive feedback cancellation. The input to the hearing-aid processing is $s(n)$, which is the sum of desired input signal $x(n)$ and the feedback signal $f(n)$. The processed hearing-aid signal is $u(n)$. The signal output into the ear canal is $y(n)$. The impulse response of the feedback path is $b(n)$, and $v(n)$ is the estimate of $f(n)$ from the modeled feedback path $\hat{b}(n,\theta)$.

shown later in Fig. 5. It is assumed in the following that both $b(n)$ and $\hat{b}(n,\theta)$ are truncated to a sufficient length so that the energy loss in the impulse response due to the truncation is at least 35 dB below the total energy of $b(n)$ or $\hat{b}(n,\theta)$.

To investigate each feedback model's capability in capturing the true feedback path, the parameters θ for each model is optimized by minimizing the difference between the modeled feedback $\hat{b}(n,\theta)$ and the actual feedback path. This minimization formulated in the frequency domain is as follows:

$$\theta_{\text{opt}} = \arg\min_{\theta} \|\mathbf{F}^H(\hat{\mathbf{b}}(\theta) - \mathbf{b})\|_2^2, \tag{1}$$

$$\mathbf{b} = [b(0), \ldots, b(L-1)]^T, \tag{2}$$

$$\hat{\mathbf{b}}(\theta) = [\hat{b}(0,\theta), \ldots, \hat{b}(L-1,\theta)]^T, \tag{3}$$

$$\mathbf{F} = [\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{L-1}] \tag{4}$$

$$\mathbf{f}_k = [1, e^{j\omega_k}, \ldots, e^{j\omega_k(L-1)}]^T, \tag{5}$$

where $\|\ \|_2$ denotes the Euclidean norm, L is the length of the truncated impulse responses in samples, $\omega_k = 2\pi k/L$, $k = 0, 1, \ldots, L-1$, $\mathbf{F}$ is the discrete Fourier transform matrix, and the superscripts T and H denote the transpose and conjugate transpose of a matrix or vector, respectively.

Based on the framework above, Secs. II A–II C will detail each feedback path model by specifying its parameter structure θ, the corresponding optimization procedure, and the analytic optimal solution if it exists. The optimal solution, either in analytic or in numerical form, represents the capability of each feedback model to represent the actual feedback path $b(n)$.

A. Existing models for dynamic feedback paths

The existing two models for dynamic feedback path are the direct model and the initialization model, both of which can be formulated as a cascade of prefiltering and an adaptive FIR filter.

Let $b_0(n)$ and $w(n)$ denote the impulse response of the prefiltering and the adaptive filter, respectively. The modeled feedback $\hat{b}(n,\theta)$ is then the convolution of $b_0(n)$ and $w(n)$, i.e.,

$$\hat{b}(n,\theta) = w(n) \odot b_0(n) = \sum_{l=0}^{M-1} w(l) b_0(n-l), \tag{6}$$

where M is the order of $w(n)$, and $\odot$ is the convolution operator.

1. Direct model

The feedback path $b(n)$ is always associated with a leading delay, which is the sum of the processing delay in $h(n)$ and the acoustic delay in the transmission path. In the "direct model," the FIR filter $w(n)$ is used to model the "active" portion of the feedback path, whereas the prefiltering $b_0(n)$ is simply a delay of D samples introduced to match the leading delay in $b(n)$ and provide a more efficient use of the limited number of taps in $w(n)$. The parameters are therefore of the form

$$b_0(n) = \begin{cases} 1, & n = D \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

$$\theta = \{w(n), D\}. \tag{8}$$

The optimal parameters $w_{\text{opt}}(n)$ and D_{opt} can be obtained by solving the optimization problem in Eqs. (1)–(8).

If the delay D is fixed, the least-squares solution is straightforward:

$$\mathbf{w}_{\text{opt}} = [b(D), \ldots, b(D+M-1)]^T. \tag{9}$$

If D is not fixed, an easy procedure to obtain the optimal solution is to search for an integer delay D that minimize the distance between $\hat{b}(n,\theta)$ and $b(n)$.

2. Initialization model

To model the feedback path accurately, the direct model in Sec. II A 1 usually needs a very high-order adaptive FIR filter $w(n)$ to cover the active range of the feedback path. One way to improve the modeling efficiency is to use the initialization model.

The feedback path $b(n)$ consists of slowly varying portions such as the responses of microphone, amplifier, receiver, etc. The idea of the initialization model is to use a fixed model, which is initialized during an additional off-line initialization procedure in the hearing-aid fitting, to model these slowly varying portions, and use a short adaptive FIR filter to represent the rapidly varying portions (mainly the change in fitting and external acoustics).

In such an initialization procedure, a "nominal feedback path" is first measured, which is defined as the feedback path without any external reflectors or enclosures near the hearing aid. When the hearing aid is put into use in daily life, the adaptive filter $w(n)$ only needs to model the part that is different from the nominal initialization to capture the time-varying dynamic feedback path. Since the impulse responses of microphone, receiver, etc., are nearly invariant from the initialization to dynamic situations, this different part is mainly resulted from the change in the hearing-aid fitting and external acoustic environments,[2] which can be modeled by an adaptive filter with a lower order.

In the initialization model, the prefiltering $b_0(n)$ in Eq. (6) is a fixed model to represent the impulse response of the nominal feedback path obtained in the initialization. In practice, it can be realized by an autoregressive moving average model (ARMA) model.[3] To avoid the complicated issue of how to model the initialized feedback path with an ARMA model to get the best generalized performance in feedback cancellation,[3,17] it is simply assumed that $b_0(n)$ is exactly the impulse response of the nominal feedback path. Therefore, the parameter for initialization model is of the form

$$\theta = \{w(n)\}. \tag{10}$$

Assume that the filter length of $w(n)$ is M, and the impulse response $b_0(n)$ is truncated to $L-M+1$ samples and padded with $M-1$ zeros so that the length of $\hat{b}(n,\theta)$, which is the convolution of $w(n)$ and $b_0(n)$, equals L. The impulse response of the dynamic feedback path $b(n)$ is also assumed to be of length L.

The optimal parameter $w_{\text{opt}}(n)$ for real valued $w(n)$ with length M in the initialization model can be found by solving the least-squares problem with Eqs. (1)–(6) and (10):

$$\mathbf{w}_{\text{opt}} = (\text{diag}(\mathbf{F}^H\tilde{\mathbf{b}}_0)\tilde{\mathbf{F}}^H)^+(\mathbf{F}^H\mathbf{b}), \tag{11}$$

$$\tilde{\mathbf{b}}_0 = [\mathbf{b}_0^T, \mathbf{0}_{1\times(M-1)}]^T, \tag{12}$$

$$\mathbf{b}_0 = [b_0(0), \ldots, b_0(L-M)]^T, \tag{13}$$

$$\tilde{\mathbf{F}} = [\tilde{\mathbf{f}}_0, \tilde{\mathbf{f}}_1, \ldots, \tilde{\mathbf{f}}_{L-1}], \tag{14}$$

$$\tilde{\mathbf{f}}_k = [1, e^{j\omega_k}, \ldots, e^{j\omega_k(M-1)}]^T, \tag{15}$$

where $\omega_k = 2\pi k/L$, $k=0,1,\ldots,L-1$, $\text{diag}(\cdot)$ forms a diagonal matrix with diagonal elements specified in $(\cdot)$, $(\cdot)^+$ is a pseudoinverse defined as $(\cdot)^+ = ((\cdot)^H(\cdot))^{-1}(\cdot)^H$, and $\mathbf{0}_{1\times(M-1)}$ represents a zero vector of size $1\times(M-1)$. The matrix $\tilde{\mathbf{F}}$ is a partial discrete Fourier transform matrix which results from the fact that $w(n)$ is only of the length M.

B. Reflection model for dynamic feedback paths

1. Model formulation

The initialization model $\hat{b}(n,w(n))$ formulated in Eq. (6) can also be regarded as a weighted sum of the nominal impulse response $b_0(n)$ and its delayed replicas with integer delays.

In typical dynamic situations where there is room reverberation or a handset placed close to the hearing aid, the feedback path is a composition of reflection and refraction, where reflection usually dominates. In addition to attenuation or amplification in the amplitude, a certain delay is associated with each reflected component. These physical delays, when transformed into the digital domain, are possibly fractional in terms of samples. Therefore, the model in Eq. (6) is generalized to a new model as follows:

$$\hat{b}(n,\theta) = \sum_{l=0}^{M-1} w(l)b_0(n-d_l), \tag{16}$$

$$\theta = \{w(l), d_l\}, \tag{17}$$

where d_l is the delay of the lth replica, $d_l > d_{l-1} \geq 0$, and $l = 1,\ldots,M-1$. Although the delays d_l in the model equations (16) and (17) are allowed to be fractional, the actual subsample implementation needs time-domain interpolation or frequency-domain multiplication. The latter is more efficient and is given in Sec. II B 2.

In this model, the impulse response of the nominal feedback path, measured during initialization without significant disturbances such as reflections, is regarded as an approximation of the direct path. The dynamic feedback path is modeled as a sum of reflection components with delay d_l and gain $w(l)$. This model is thus named "reflection model." When $d_l = l$, the reflection model reduces to the initialization model. Therefore, the reflection model is more general and expected to capture the dynamic feedback path better than the initialization model since it represents more accurately what happens in the physical world.

2. Delay estimation

The optimal delays $d_{l,\text{opt}}$ and weights $w_{\text{opt}}(l)$ for the reflection model can be found by solving the optimization problem given by Eqs. (1)–(5), (16), and (17), which is a nonlinear optimization problem.

However, efficient time delay (TDE) and amplitude estimation techniques have been investigated intensively in many fields such as radar, sonar, radio navigation, geophysical/seismic exploration, wireless communication, and medical imaging. The most well-known approach is

based on the matched filter,[18] whose resolution capability is unfortunately limited to the reciprocal of the signal bandwidth. For dynamic feedback path modeling, the difference between delays can be very small due to the handset proximity effect. Therefore, TDE techniques with high resolution should be used. Among these techniques, algorithms such as multiple signal classification, linear prediction, and maximum likelihood are not very well suited here since they are best for complex-valued signals with special spectral shapes.[19] A weighted Fourier transformation and relaxation (WRELAX) based method[20] is found to be very robust to address the problem. First, the cost function is defined as follows:

$$C_1(\{w(l),d_l\}_{l=0}^{M-1}) = \left\| \mathbf{B} - \sum_{l=0}^{M-1} w(l)[\mathbf{S}\mathbf{a}(d_l)] \right\|_2^2, \tag{18}$$

where

$$\mathbf{B} = \mathbf{F}^H \mathbf{b}, \tag{19}$$

$$\mathbf{S} = \mathrm{diag}(\mathbf{F}^H \tilde{\mathbf{b}}_0), \tag{20}$$

$$\mathbf{a}(d_l) = [1, e^{-j(2\pi/L)d_l}, \ldots, e^{-j(2\pi/L)d_l(L-1)}]^T. \tag{21}$$

Equation (18) is simply a formulation of the inner part of Eq. (1) with Eqs. (16) and (17) inserted. Minimizing the cost function $C_1(\{w(l),d_l\}_{l=0}^{M-1})$ with respect to $w(l)$ and d_l is the problem of interest. To address this optimization problem, denote

$$\mathbf{B}_l = \mathbf{B} - \sum_{i=0,i\neq l}^{M-1} \hat{w}(i)[\mathbf{S}\mathbf{a}(\hat{d}_i)], \tag{22}$$

$$C_2(w(l),d_l) = \|\mathbf{B}_l - w(l)\mathbf{S}\mathbf{a}(d_l)\|_2^2. \tag{23}$$

The cost function $C_2(w(l),d_l)$ assumes that the other reflection components have been estimated as $\hat{w}(i)$ and $\hat{d}_i$, where $i=0,1,\ldots,M-1, i\neq l$, except the lth component, and minimizing this cost function yields the estimates $\hat{w}(l)$ of $w(l)$ and $\hat{d}_l$ of d_l as[20]

$$\hat{d}_l = \arg\max_{d_l} |\mathbf{a}^H(d_l)(\mathbf{S}^*\mathbf{B}_l)|^2, \tag{24}$$

$$\hat{w}(l) = \left. \frac{\mathbf{a}^H(d_l)(\mathbf{S}^*\mathbf{B}_l)}{\|\mathbf{S}\|_2^2} \right|_{d_l=\hat{d}_l}, \tag{25}$$

where the asterisk denotes the complex conjugate. If $w(l)_{l=0}^{M-1}$ are assumed to be real numbers, minimizing $C_2(w(l),d_l)$ with respect to $w(l)$ and d_l yields[20]

$$\hat{d}_l = \arg\max_{d_l} \{\Re[\mathbf{a}^H(d_l)(\mathbf{S}^*\mathbf{B}_l)]\}^2, \tag{26}$$

$$\hat{w}(l) = \left. \frac{\Re[\mathbf{a}^H(d_l)(\mathbf{S}^*\mathbf{B}_l)]}{\|\mathbf{S}\|_2^2} \right|_{d_l=\hat{d}_l}, \tag{27}$$

where $\Re[\cdot]$ denotes the real part of $(\cdot)$.

To estimate $\hat{d}_l$ using Eq. (26), an initial estimate of d_l is first obtained by locating the maximum peak in the magnitude of the K-point ($K \geq L$) fast Fourier transform (FFT) of $\mathbf{S}^*\mathbf{B}_l$ with padded zeros. This process can also be regarded as a matched filtering in the frequency domain. Then a search for the $\hat{d}_l$ that maximizes $\Re\{[\mathbf{a}^H(d_l)(\mathbf{S}^*\mathbf{B}_l)]\}^2$ is made in a small range around the initial estimate based on golden section search and parabolic interpolation.[21,22]

The WRELAX estimates the delays d_l and amplitudes w_l in the frequency domain to avoid fractional delay interpolation. The general idea is to cross-correlate $b(n)$ and $b_0(n)$ in the frequency domain to find the coarse delays and gains of the replicas by identifying the peaks of the cross-correlation. Later an iterative search is performed by keeping one replica of $b_0(n)$ at a time (removing the other identified replicas from $b(n)$), repeating the cross-correlation and locating the peak to find a better delay and gain estimation for that replica. This process is iterated until the relative change in the cost function in Eq. (18) is below the threshold ϵ. The detailed steps are given below.

Step 1: For $M=1$ obtain $\{\hat{w}(l),\hat{d}_l\}_{l=0}$ from $\mathbf{B}$ by using Eqs. (26) and (27).

Step 2: For $M=2$ compute $\mathbf{B}_1$ with Eq. (22) by using $\{\hat{w}(l),\hat{d}_l\}_{l=0}$ obtained in Step 1. Obtain $\{\hat{w}(l),\hat{d}_l\}_{l=1}$ from $\mathbf{B}_1$ by using Eqs. (26) and (27). Next recompute $\mathbf{B}_0$ by using $\{\hat{w}(l),\hat{d}_l\}_{l=1}$ and redetermine $\{\hat{w}(l),\hat{d}_l\}_{l=0}$ from $\mathbf{B}_0$. Iterate these two substeps until the relative change in $C_1(\{w(l),d_l\}_{l=0}^{M-1})$ between two consecutive iterations is below the threshold ϵ.

Step 3: For $M=3$ compute $\mathbf{B}_2$ with Eq. (22) by using $\{\hat{w}(l),\hat{d}_l\}_{l=0}^{1}$ obtained in Step 2. Obtain $\{\hat{w}(l),\hat{d}_l\}_{l=2}$ from $\mathbf{B}_2$. Next recompute $\mathbf{B}_0$ by using $\{\hat{w}(l),\hat{d}_l\}_{l=1}^{2}$ and redetermine $\{\hat{w}(l),\hat{d}_l\}_{l=0}$ from $\mathbf{B}_0$. Then compute $\mathbf{B}_1$ by using $\{\hat{w}(l),\hat{d}_l\}_{l=0,2}$ and redetermine $\{\hat{w}(l),\hat{d}_l\}_{l=1}$. Iterate these three substeps until the relative change in $C_1(\{w(l),d_l\}_{l=0}^{M-1})$ between two consecutive iterations is below the threshold ϵ.

Remaining Steps: Continue similarly until M is equal to the desired number of paths.

In contrast to some estimation algorithms, such as the expectation-maximization algorithm,[23] the WRELAX is not sensitive to the initial parameters since it uses a matched filter to find these initial values. It is bound to converge to at least a local minimum point, which is at least a better solution than what the traditional matched filter can yield. The convergence speed of WRELAX depends on the time delay spacing of the different reflected paths.[20]

C. Models for dual-microphone hearing aids

For hearing aids with dual microphones, the feedback problem involves a dynamic feedback path for the front microphone, denoted as $b_f(n)$, and a path for the rear microphone, denoted as $b_r(n)$. One way of dealing with the two paths is to model them individually by using one of the three models described above. An alternative approach involves

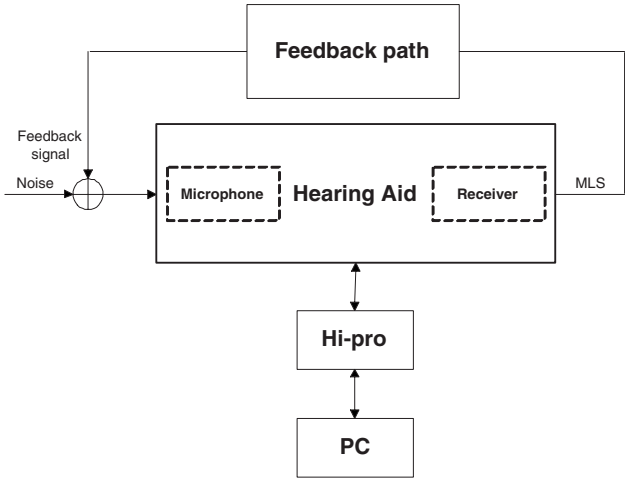


FIG. 2. The block diagram of the measurement system for measuring the feedback paths. The feedback path is illustrated which includes the external acoustics, for example, a Kemar head. The receiver sends out a MLS and the feedback signal is picked up by the microphone with ambient noise together. The data are fed into the computer through HI-PRO.

exploiting the relation between the two similar feedback paths by fitting one dynamic feedback path with the other dynamic path. There are two ways of fitting.

The first approach for the fitting, which is similar to the initialization model [cf. Eq. (6)], is

$$\hat{b}_r(n) = \sum_{l=0}^{M-1} w(l) b_f(n-l). \tag{28}$$

The second approach for the fitting, which is similar to the reflection model [cf. Eq. (16)], is

$$\hat{b}_r(n) = \sum_{l=0}^{M-1} w(l) b_f(n-d_l). \tag{29}$$

The estimate of $w_{\text{opt}}(n)$ and/or delays $d_{l,\text{opt}}$ can be found in similar ways described in Secs. II A 2 and II B.

III. MEASUREMENT OF DYNAMIC FEEDBACK PATHS

The nominal and dynamic feedback paths have been measured using a commercial open-fitting BTE device with two microphones from GN ReSound A/S. The hearing aid was mounted on the head of Kemar Manikin type 45BA made by G.R.A.S. Sound & Vibration A/S.[24]

The impulse responses of the feedback paths were measured by sending out a maximum-length sequence (MLS) with a period of 255 samples through the receiver, repeating it 1000 periods, and averaging the responses to get a high signal-to-noise ratio (SNR) for the feedback path response relative to random ambient noise. The sampling frequency was 15625 Hz. Figure 2 illustrates the measurement diagram where a HI-PRO universal programming interface for hearing instruments made by GN Otometrics A/S (Ref. 25) is used to feed the collected data into a computer to calculate the impulse responses. The detailed procedure of the impulse response measurement can also be found in Ref. 10.

The MLS is emitted at around 85 dB sound pressure level (SPL) and transmitted to the microphone through the feedback. The feedback path shapes the signal based on its frequency response. The internal noise in the microphone is less than 28 dB SPL and the ambient noise is less than 30 dB SPL. The average over 1000 cycles of MLS increases the overall SNR by 30 dB. This setup has been found to be very reliable. The preliminary test shows that when the measurements are repeated for ten times for a single feedback path, the variance in the measured impulse responses is around 36 dB below the average response, which is sufficient for feedback path modeling.

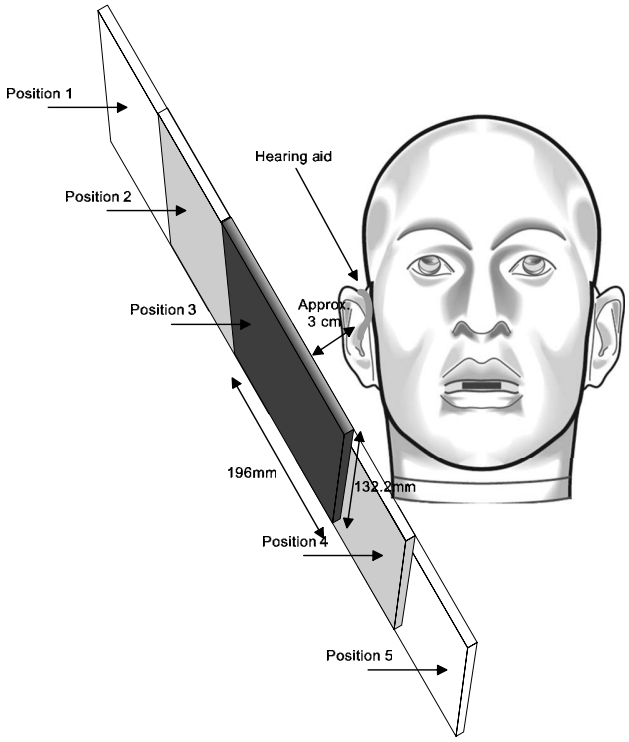


FIG. 3. The measurement setup for dynamic feedback paths with rigid surface. The hearing aid is mounted on the Kemar head. Positions 1 and 2 are symmetrical to positions 4 and 5 with regard to position 3, which is right at the side of the hearing aid.

The nominal feedback paths for both the front and rear microphones are measured. The measured nominal feedback path for the front microphone is denoted as $b_{f,0}(n)$ and for the rear microphone $b_{r,0}(n)$.

Since it was found that the movement of jaw, mouth, etc., had little effect on the feedback path variations,[5,2] and the adaptive feedback cancellation in hearing aids today usually handles these situations very well in practice, the measurement of dynamic feedback paths focused on the proximity effect.

Seven dynamic feedback paths were measured for both the front and rear microphones. A rigid metal surface of dimension 196×132.6 mm^2 was used and faced to the hearing aid mounted on the Kemar head. The surface was moved along the lateral side gradually toward the hearing aid from position 1 to position 3 and later from position 3 to position 5, as shown in Fig. 3. The perpendicular distance between the rigid surface and the hearing aid was kept at around 3 cm during the movement. The impulse responses of the dynamic feedback path at these five positions measured from the front

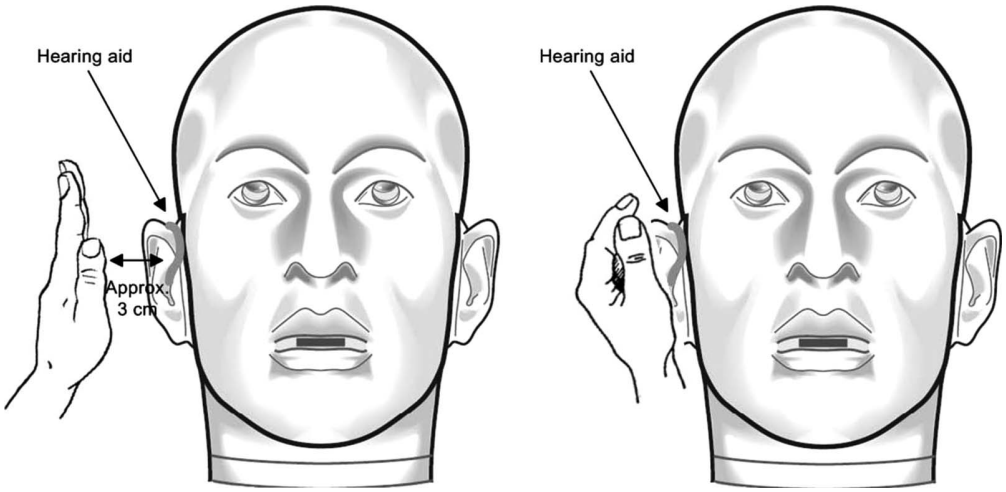


FIG. 4. The measurement setup for dynamic feedback paths with palm. The left figure represents an outstretched palm facing the hearing aid, and the right figure depicts the cupped hand around the hearing aid.

microphone are denoted as $b_{f,1}(n)-b_{f,5}(n)$ and from the rear microphone $b_{r,1}(n)-b_{r,5}(n)$. Apart from the rigid surface, two additional dynamic feedback paths were measured with an outstretched palm facing the hearing aid on its lateral side at a distance of around 3 cm $[b_{f,6}(n),b_{r,6}(n)]$ and with a palm wrapping around the hearing aid $[b_{f,7}(n),b_{r,7}(n)]$, as illustrated in Fig. 4.

Therefore, altogether eight impulse responses were measured for both front and rear microphones including the static feedback path (nominal feedback) and seven dynamic feedback paths. All the impulse responses are of the length 255 samples. The rigid surface measurements are artificial tests for reflections whereas the palm measurements mimic the most adverse situations for feedback cancellation in real life.

The frequency responses of these feedback paths are shown in Fig. 5, where only the responses from the front microphone are illustrated since the responses from the rear are similar.

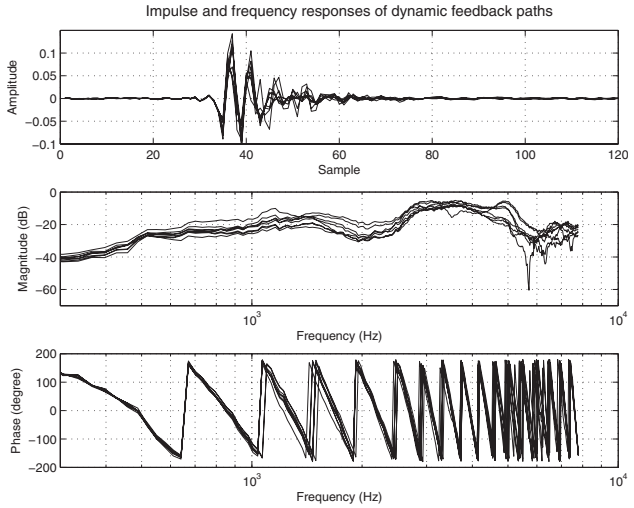


FIG. 5. Impulse responses and frequency responses of measured eight feedback paths (one nominal and seven dynamic feedback paths).

IV. PROPERTIES OF DYNAMIC FEEDBACK PATHS

In this section, the general characteristics of the measured dynamic feedback paths are investigated first. Later, the reflection model proposed in Sec. II B is used to model these paths and reveal other aspects of the characteristics of dynamic feedback paths.

A. Variability of dynamic feedback paths

The sample mean and sample estimate of the standard deviation of the measured dynamic feedback paths have been calculated as a function of frequency. The mean and standard deviation do not reveal all the properties of the dynamic feedback paths but give an indication of the dynamic range of the data. If it can be assumed that the variation in spectrum magnitude for each frequency is normally distributed, then the 95% confidence interval of the expected value for seven measurements is $[\bar{\mathbf{B}}_f-0.925\mathbf{S}_f,\bar{\mathbf{B}}_f+0.925\mathbf{S}_f]$ and $[\bar{\mathbf{B}}_r-0.925\mathbf{S}_r,\bar{\mathbf{B}}_r+0.925\mathbf{S}_r]$ for the front- and rear-microphone dynamic paths, respectively, where $\bar{\mathbf{B}}_f$ and $\bar{\mathbf{B}}_r$ are sample means, and $\mathbf{S}_f$ and $\mathbf{S}_r$ are sample estimates of the standard deviation.

They are calculated as follows:

$$\bar{\mathbf{B}}_f=\frac{1}{7}\sum_{l=1}^{7}|\mathbf{F}^H\mathbf{b}_{f,l}|, \tag{30}$$

$$\mathbf{S}_f=\sqrt{\frac{1}{7-1}\sum_{l=1}^{7}(|\mathbf{F}^H\mathbf{b}_{f,l}|-\bar{\mathbf{B}}_f)^2}, \tag{31}$$

where

$$\mathbf{b}_{f,l}=[b_{f,l}(0),\ldots,b_{f,l}(L-1)]^T. \tag{32}$$

Note that $\bar{\mathbf{B}}_f$ and $\mathbf{S}_f$ are both vectors. The symbol $|\cdot|$ in Eq. (30) and $(\cdot)^2$ in Eq. (31) denote the elementwise absolute operation and elementwise square of the vector in $(\cdot)$, respectively. The $\bar{\mathbf{B}}_r$ and $\mathbf{S}_r$ can be calculated similarly.

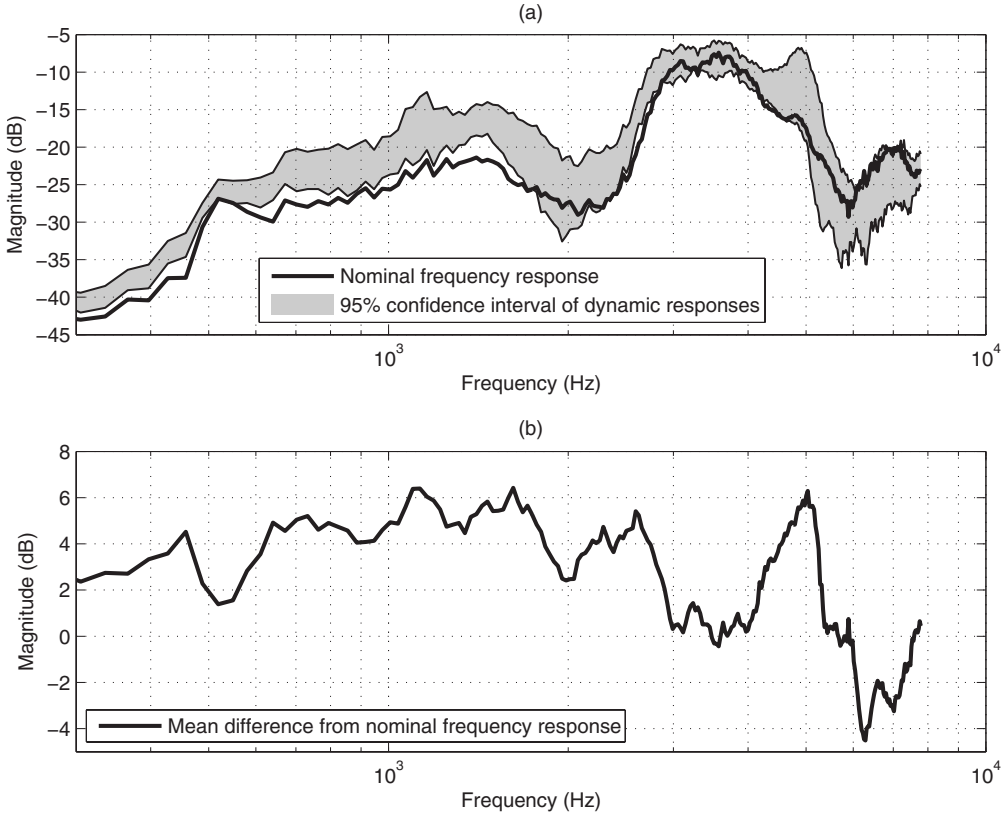


FIG. 6. Variability of dynamic feedback paths: (a) shows the nominal frequency response and a shaded region representing the 95% confidence interval of the expected frequency response of dynamic feedback paths; (b) illustrates the mean difference between the responses of dynamic feedback paths and the nominal feedback path.

The variability results are shown in Fig. 6(a). The shaded region represents the uncertainty in the expected magnitude of frequency responses of the dynamic feedback paths. It should be noted that the variability illustrated here is obtained with one single head and does not include the intersubject variability. The general trend is that at low frequencies, the uncertainty is small, but above 1 kHz, the uncertainty increases and goes up to 10 dB at some frequencies. The magnitude curve of the frequency response of the nominal feedback path lies below this shaded region at frequencies lower than 1750 Hz, which means that the rigid surface or hand in the measurements generally boosted the feedback responses at low frequencies. Above 1750 Hz, most of the nominal curve lies within the shaded region, and above 5 kHz, nearly all of the curve lies inside the region. This indicates that the change in feedback path at high frequencies is not a general boost but resulted from complicated effects such as peaks and valleys. This trend agrees with the results from other studies,[5,2,6] despite the minor differences in the measured feedback paths due to different devices, test environments, etc.

The difference in dB between the sample mean of the magnitude of dynamic feedback paths $\bar{\mathbf{B}}_f$ and the magnitude of the nominal feedback path is illustrated in Fig. 6(b), which essentially removes the effects of microphone, receiver, etc. A dominant peak was found at 5035 Hz. This can be interpreted as a strong reflection from an object placed at a distance of $1/5035 \times 343/2$ m ≈ 0.034 m, where 343 m/s is the speed of sound at 20 °C and the factor of 2 is used to calculate a single-way transmission length. This is approximately the distance between the hand/rigid surface and the hearing aid in the measurement.

B. Reflections in dynamic feedback paths

The reflection phenomenon in dynamic feedback paths with a telephone handset near the hearing aid has been pointed out in several previous papers.[3,5,6] It features peaks or other complicated effects in the high frequency range of the measured feedback responses. The dominant peak in the Fig. 6(b) also indicates the occurrence of a reflection.

However, physically, there could be many other reflections besides the dominant one, especially when the object is placed very close to the hearing aid. The reflection model proposed in this paper can be used to estimate these reflections.

In WRELAX of the reflection model, one zero is padded to all of the impulse responses of the measured feedback paths so that the truncated length L becomes 256. The number of points in the FFT for initial delay estimation, K, is set to 4096 to get a good initial guess, the search region around the initial estimate $\hat{d}_{l,\mathrm{ini}}$ in samples is set to be $[\hat{d}_{l,\mathrm{ini}} - \frac{1}{32}, \hat{d}_{l,\mathrm{ini}} + \frac{1}{32}]$, and the threshold for the stopping criterion ϵ is set as 10^{-4}.

The reflection model estimates the $w(l)$ and d_l, where $l = 0, 1, \ldots, M-1$, for all the measured data. In Figs.

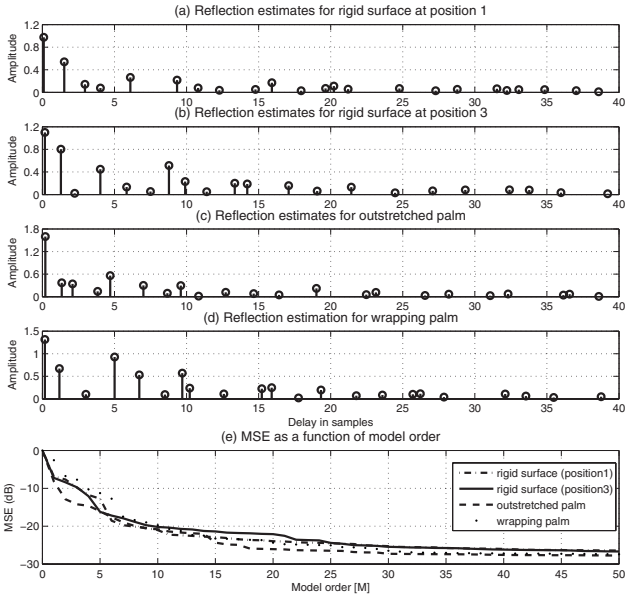


FIG. 7. (a)–(d) illustrate the reflection estimates in terms of delay in samples $\hat{d}_l$ and amplitude $\hat{w}(l)$ based on the measured data for rigid surface at positions 1 and 3, outstretched palm, and wrapping palm; (e) shows how the MSE, i.e., $C_1(\{w(l),d_l\}_{l=0}^{M-1})$ decreases as the number of reflection components M increases.

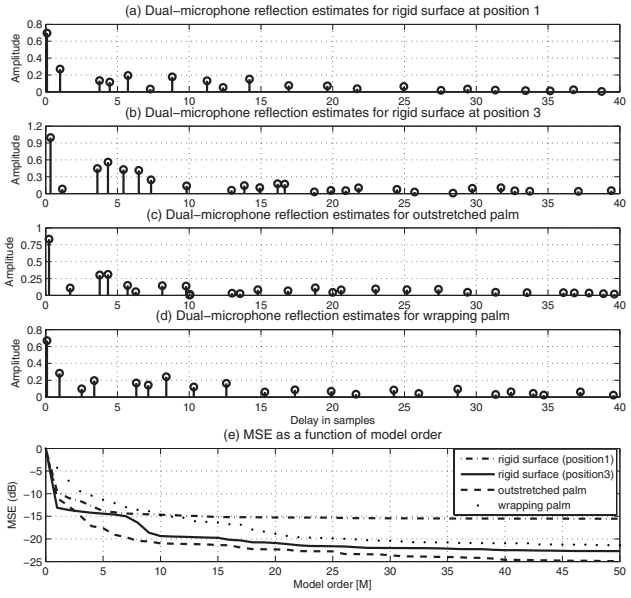


FIG. 8. (a)–(d) illustrate the reflection estimates for the dual-microphone BTE hearing aid assuming that the rear response is a sum of reflected replicas of the front response in terms of delay in samples $\hat{d}_l$ and amplitude $\hat{w}(l)$ based on the measured data for rigid surface at positions 1 and 3, outstretched palm, and wrapping palm; (e) shows how the MSE, i.e., $C_1(\{w(l),d_l\}_{L=0}^{M-1})$ decreases as the number of reflection components M increases.

7(a)–7(d), the results for the rigid surface at positions 1 and 3, and palm outstretched and wrapped around the hearing aid, are illustrated. The MSE of the estimated feedback path, i.e., $C_1(\{w(l),d_l\}_{l=0}^{M-1})$ defined in Eq. (18), is given in Fig. 7(e) as a function of the model order. As can be seen from the figure, in all the four cases, the highest peak is located near 0 sample delay, which implies that the assumption that the nominal feedback path is approximately the direct path is valid. The reason why the amplitude is not exactly unity and the delay is not exactly 0 is that the nominal feedback path itself consists of a number of small reflections due to the presence of pinna and head.

As the delay increases, the estimated amplitude generally decays, which complies with the physical law that longer transmission path results in larger attenuation (transmission loss).[26] Almost all of the significant peaks in the four cases happen before a delay of ten samples, corresponding to a single-way transmission length of 0.106 m, which is a reasonable range for late reflection paths, e.g., a path consisting of multiple reflections between reflecting objects, the ear, and the hearing aid before it reaches the microphone.

There are fewer significant reflections (with amplitude larger than 0.4) in the cases of the rigid surface at position 1 and outstretched palm than in the cases of the rigid surface at position 3 and wrapping hand. This agrees with the fact that the reflections in the latter two cases are more complicated.

The case with the wrapping hand has the most irregular reflection pattern, in which large reflections are distributed widely in the delay range from zero to ten samples. The reflections for this situation after a delay of ten samples are generally as small as in the other cases. This shows that the wrapping hand introduces large reflections with short delays.

In the experiments with the rigid surface and outstretched palm, the distance was kept at around 3 cm, corresponding to a delay of 2.73 samples and a peak at 5717 Hz in the magnitude response. The peak has been verified in Sec. IV A, whereas for all the cases in Figs. 7(a)–7(d), the peak does not seem to be very dominant, although there is a reflection component located at around 2.73 samples. A possible reason is that the dominant reflections were influenced by the presence of the pinna and the head, depending on how the reflecting object was positioned. However, this influence is different for different measurements. Therefore, it is averaged out and the dominant peak shows up at a correct location in the average magnitude response [see Fig. 6(b) at 5035 Hz].

For the dual-microphone situation, the estimation results are shown in Fig. 8. For each case, the response of the feedback path at the rear microphone is fitted by that at the front microphone using the proposed reflection model. As seen from Fig. 8(e), the MSE converges at a relatively higher level compared with Fig. 7(e), especially for the case where the rigid surface is placed at position 1. Therefore, the usage of a reflection model to relate the two feedback paths of a dual-microphone BTE hearing aid is not very effective.

V. PERFORMANCE OF THE PROPOSED MODEL

In this section, the proposed reflection model is compared with the two existing model in terms of MSE and MSG. The performance of the reflection model for dual-microphone situation is also investigated.

A. Performance metric of feedback path models

To evaluate the performance of a feedback model, two metrics are usually used: MSE and MSG. The MSE de-

scribes the average distance between the modeled and the real feedback paths. The MSG indicates the maximum gain without instability assuming a flat frequency response in the hearing-aid processing and the worst case for the phase. The MSG is determined by the frequency at which the mismatch between the feedback model and the actual feedback path is the largest.[10] Assume that all the models are optimized, the resultant MSE and MSG are defined as MSE_c and MSG_c, respectively,

$$\mathrm{MSE}_c = 20\log_{10}\left(\frac{\|\mathbf{F}^H\hat{\mathbf{b}}(\theta_{opt}) - \mathbf{b}\|_2^2}{\|\mathbf{F}^H\mathbf{b}\|_2^2}\right), \tag{33}$$

$$\mathrm{MSG}_c = 20\log_{10}\left(\min_k \frac{1}{|\mathbf{f}_k^H(\hat{\mathbf{b}}(\theta_{opt}) - \mathbf{b})|}\right). \tag{34}$$

With a specific model and parameters θ, the MSE_c and MSG_c are the lowest achievable MSE and the highest achievable MSG, respectively. MSE_c and MSG_c are, in fact, limited by the amount of undermodeled feedback path, the residual feedback path that cannot be modeled due to the limited number of degrees of freedom in the parameter θ and/or the lack of flexibility in the model form. A more descriptive model with larger degrees of freedom in the parameters θ will yield less undermodeling, lower MSE_c and larger MSG_c.

B. Results

For each measured dynamic feedback path, the parameters in the models were first optimized to calculate the MSE_c and MSG_c. The filter length M was varied from 1 to 50. To achieve the best performance, the delay D was not fixed in the direct model.

It is found that for all the seven dynamic paths and all the values M, the reflection model outperforms the initialization model and the direct model in terms of MSE_c and MSG_c. The direct model performs worst in all the cases. To demonstrate the performance of each model in dynamic situations, MSE_c and MSG_c are averaged over the seven dynamic paths. The results are illustrated in Figs. 9 and 10. The results for the dual-microphone models are also included.

As seen in the figures, the general trend of MSE_c and MSG_c is similar, the larger M the smaller MSE_c and the larger MSG_c. The slight difference in the pattern between MSE_c and MSG_c is because the optimization of parameters in the models is based on MSE instead of MSG directly.

The reflection model is superior to the other two models, especially when M is between 6 and 21. In practice, M is usually chosen between 10 and 20 to assure a fast convergence. In this region, the reflection model yields 5–6 dB higher MSE_c, 5–7 dB higher MSG_c than the initialization model, and 8–10 dB higher MSE_c, and 9–11 dB higher MSG_c than the direct model. To achieve a 25 dB MSG_c, the direct model needs 31 orders and the initialization model needs 16 orders, whereas the reflection model only needs seven reflection replicas of the nominal impulse response. Moreover, including 11 reflection replicas in the reflection model yields around the same MSG_c as the initialization model with 50-order FIR filter. This shows that 11 reflections are enough to model the dominant characteristics of the dynamic feedback paths resulted from a hand or rigid surface placed close to the hearing aid.

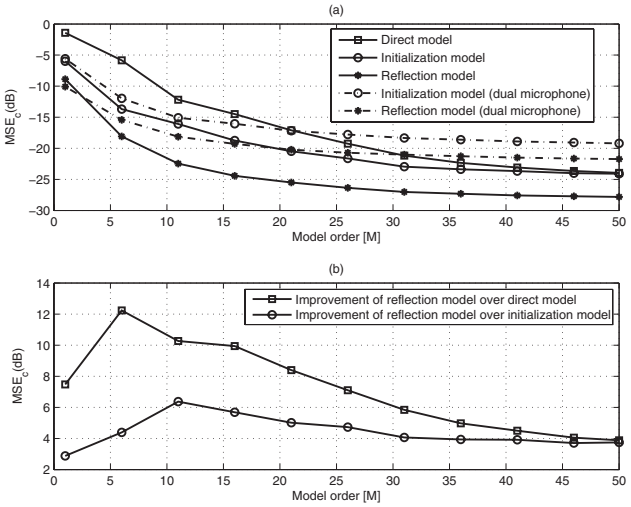


FIG. 9. Comparison of the models for dynamic feedback path modeling in terms of MSE_c. (b) shows the MSE_c improvement of the reflection model over direct model and initialization mode. The horizontal axis is the filter length in the direct model and initialization model or the number of reflection components in reflection model.

It should be noted that the reflection model does not yield significant improvement when used in the dual-microphone case. This is because in a complex acoustic environment, the relation between the two feedback paths is very complicated and even more difficult to model than the feedback paths themselves.

VI. CONCLUSIONS AND FUTURE WORK

This paper describes a novel reflection model for the dynamic feedback path in digital hearing aids. The reflection

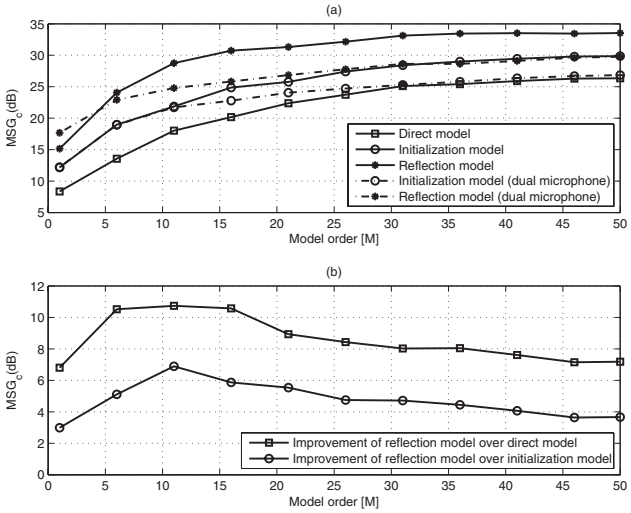


FIG. 10. Comparison of the models for dynamic feedback path modeling in terms of MSG_c. (b) shows the MSG_c improvement of the reflection model over direct model and initialization mode. The horizontal axis is the filter length in the direct model and initialization model or the number of reflection components in reflection model.

properties in addition to the variability of feedback path, when the objects such as a palm or handset are placed close to the hearing aid, have been investigated.

The results based on the measurement of a commercial hearing aid show that the proposed model has a better ability of capturing the dynamic feedback path in these difficult real-life situations and is superior to the existing two models in terms of MSE and MSG. To achieve the same level of MSE or MSG, the number of components required in the proposed model is fewer than the orders of the FIR filter required in the two traditional models.

The results also show the minimum order of the adaptive filter in the two existing models to achieve a certain MSG in the dynamic situations, which could serve as a useful indication in practice for choosing the order of the adaptive filter in feedback cancellation. For the reflection model, 11 reflection replicas are enough to model the dominant characteristics of the proximity effect in dynamic feedback paths.

This paper also investigates the possibility of relating the two feedback paths of a dual-microphone hearing aid for modeling the dynamic feedback paths. It is shown that in a complicated acoustic environment, the relation between the two feedback paths can be very intricate and difficult to exploit to yield better modeling of the dynamic feedback paths.

The future work is to develop an on-line adaptation scheme for this reflection model so that it can improve the performance of feedback cancellation in adverse dynamic situations. The first step to this adaptation scheme is to find a more efficient way of estimating the fractional delays. The frequency-domain approach will be preferred since the implementation of a fractional delay in the time domain requires interpolation. The next step is to find a proper frequency-domain adaptive filtering approach to estimate the dynamic feedback paths based on the estimated delays.

[1]S. F. Lybarger, "Acoustic feedback control," in *The Vanderbilt Hearing-Aid Report*, edited by Studebaker and Bess (Monographs in Contemporary Audiology, Upper Darby, PA, 1982), pp.87–90.

[2]B. Rafaely, M. Roccasalva-Firenze, and E. Payne, "Feedback path variability modeling for robust hearing aids," J. Acoust. Soc. Am. **107**, 2665–2673 (2000).

[3]J. Kates, "Constrained adaptation for feedback cancellation in hearing aids," J. Acoust. Soc. Am. **106**, 1010–1019 (1999).

[4]J. Kates, *Digital Hearing Aids* (Plural, San Diego, CA, 2008).

[5]J. Hellgren, T. Lunner, and S. Arlinger, "Variations in the feedback of hearing aids," J. Acoust. Soc. Am. **106**, 2821–2833 (1999).

[6]M. Stinson and G. Daigle, "Effect of handset proximity on hearing aid feedback," J. Acoust. Soc. Am. **115**, 1147–1156 (2004).

[7]D. P. Egolf, B. T. Haley, K. A. Weaver, and D. S. Barker, "The hearing aid feedback path: Mathematical simulations and experimental verification," J. Acoust. Soc. Am. **78**, 1578–1587 (1985).

[8]D. P. Egolf, B. T. Haley, H. C. Howell, and S. Legowski, "Simulating the open-loop transfer function as a means for understanding acoustic feedback in hearing aids," J. Acoust. Soc. Am. **85**, 454–467 (1989).

[9]K. Nakao, R. Nishimura, and Y. Suzuki, "Calculation of transfer function of acoustic feedback path for in-the-ear hearing aids with a correction for specific acoustic impedance of a tubule," Acoust. Sci. & Tech. **27**, 242–244 (2006).

[10]J. Kates, "Room reverberation effects in hearing aid feedback cancellation," J. Acoust. Soc. Am. **109**, 367–378 (2001).

[11]J. Maxwell and P. Zurek, "Reducing acoustic feedback in hearing aids," IEEE Trans. Speech Audio Process. **3**, 304–313 (1995).

[12]J. Kates, "Feedback cancellation in hearing aids: A computer simulation results," IEEE Trans. Signal Process. **39**, 553–562 (1991).

[13]M. M. A. Spriet, I. Proudler, and J. Wouters, "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," IEEE Trans. Signal Process. **53**, 3749–3763 (2005).

[14]J. Hellgren, "Analysis of feedback cancellation in hearing aids with filtered-x lms and the direct method of closed loop identification," IEEE Trans. Acoust. Speech Signal Process. **10**, 119–131 (2002).

[15]R. J. Oliveira, "The active earcanal," J. Am. Acad. Audiol. **8**, 401–410 (1997).

[16]G. Ma, F. Gran, F. Jacobsen, and F. Agerkvist, "A new approach for modelling the dynamic feedback path of digital hearing aids," in 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (2009), pp. 209–212.

[17]Y. Jingbo, T. Mengtong, and J. Chang, "Modeling external feedback path of an itedigital hearing instrument for acoustic feedback cancellation," in 2005 IEEE International Symposium on Circuits and Systems (2005), pp. 1326–1329.

[18]J. Ehrenberg, T. Ewart, and R. Morris, "Signal-processing techniques for resolving individual pulses in a multipath signal," J. Acoust. Soc. Am. **63**, 1861–1865 (1978).

[19]R. Wu, J. Li, and Z.-S. Liu, "Super resolution time delay estimation via mode-wrelax," IEEE Trans. Aerosp. Electron. Syst. **35**, 294–307 (1999).

[20]J. Li and R. Wu, "An efficient algorithm for time delay estimation," IEEE Trans. Signal Process. **46**, 2231–2235 (1998).

[21]R. Brent, *Algorithms for Minimization Without Derivatives* (Prentice-Hall, Englewood Cliffs, NJ, 1973).

[22]G. Forsythe, M. Malcolm, and C. Moler, *Computer Methods for Mathematical Computations* (Prentice-Hall, Englewood Cliffs, NJ, 1976).

[23]M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," IEEE Trans. Acoust. Speech Signal Process. **36**, 477–489 (1988).

[24]Product Data and Specifications of Kemar Manikin Type 45BA, G.R.A.S. Sound & Vibration, Holte, Denmark (2007).

[25]HI-PRO Hardware Installation Guide, GN Otometrics A/S, Taastrup, Denmark (1993).

[26]L. L. Beranek, *Acoustics* (McGraw-Hill, New York, 1954).

Paper II

Adaptive Feedback Cancellation with Band-limited LPC Vocoder in Digital Hearing Aids

G. Ma, F. Gran, F. Jacobsen and F. T. Agerkvist

Reviewed, revised and resubmitted to IEEE Transactions on Audio, Speech and Language Processing, March 2010

Adaptive feedback cancellation with band-limited LPC vocoder in digital hearing aids

Guilin Ma, *Student Member, IEEE*, Fredrik Gran, Finn Jacobsen and Finn T. Agerkvist

***Abstract*—Feedback oscillation is one of the major issues with hearing aids. An effective way of feedback suppression is adaptive feedback cancellation, which uses an adaptive filter to estimate the feedback path. However, when the external input signal is correlated with the receiver input signal, the estimate of the feedback path is biased. This so-called "bias problem" results in a large modeling error and a cancellation of the desired signal. This paper proposes a band-limited linear predictive coding based approach to reduce the bias. The idea is to replace the hearing-aid output with a synthesized signal, which sounds perceptually the same as or similar to the original signal but is statistically uncorrelated with the external input signal at high frequencies where feedback oscillation usually occurs. Simulation results show that the proposed algorithm can effectively reduce the bias and the misalignment between the real and the estimated feedback path. When combined with filtered-X adaptation in the feedback canceller, this approach reduces the misalignment even further.**

I. Introduction

Feedback in a hearing aid refers to a process in which some of the receiver output is picked up by the microphone. The acoustic feedback path is the most significant contributor to the feedback signal although electrical and mechanical paths also exist [1]. A typical acoustic feedback path of the hearing aid represents a wave propagation path from the receiver to the microphone, which includes the effects of the hearing-aid receiver, microphone, the acoustics of the vent or leak, etc. The hearing-aid processing amplifies the input signal to compensate for the hearing loss of the users. When this amplification is larger than the attenuation of the feedback path, instability occurs and usually results in feedback whistling, which limits the maximum gain that can be achieved [2] and compromises the comfort of wearing hearing aids.

A widely adopted approach to acoustic feedback suppression is adaptive feedback cancellation (AFC), in which the acoustic feedback signal is estimated by an adaptive filter and then subtracted from the input signal to remove feedback [3]. A perfect match between the modeled and the real feedback path will cancel the feedback signal completely, and prevent instability for any amount of amplification. However, in practice, there is always a modeling error for many reasons, such as too slow adaptation speed, insufficient filter length, etc. A significant portion of the modeling error is the result of a so-called "bias problem", which refers to a biased estimate of the feedback path when the desired input signal and the the receiver input signal are correlated [4]. During the past two decades, various approaches have been proposed to decorrelate the input and output of a hearing aid to reduce the bias in the estimate of the feedback path.

One well-known decorrelation approach introduces a delay in the hearing-aid processing (or the feedback cancellation path) to decorrelate the input of the receiver and the incoming signal. It has been shown in [4] that for a colored noise input, the insertion of delay in the hearing-aid processing significantly improves the accuracy of feedback modeling, while the insertion of a delay in the feedback cancellation path provides smaller benefit. However, the delay introduced in the hearing aids should be kept small to avoid disturbing artifacts such as comb filtering [5]. Moreover, for tonal signals, a delay will not help much to reduce the correlation.

Another kind of decorrelation approach uses nonlinearities in the hearing-aid processing. Methods based on this approach include frequency shifting [6], phase or delay modulation [7], time-varying all-pass filter [8], etc. Some of these methods can only yield limited improvement. For example, the delay and phase modulation improve the gain before instability by 2 dB and 4 dB, respectively [7]. Since all the nonlinear methods degrade sound quality to some extent, a trade-off between the performance of feedback cancellation and sound quality is usually involved.

Alternatively, a probe signal, usually a noise signal, is added to the receiver input [9]. To maintain sound quality, the probe signal should be inaudible and its level therefore has to be much lower than that of the original receiver input signal. The bias reduction achieved with such a weak probe signal is very small.

A recently proposed decorrelation method exploits closed-loop identification techniques and uses linear prediction in the feedback cancellation path [10–12]. In [12] it has been proven that by minimizing the prediction error of the microphone signal, the estimate of the feedback path is not biased (identifiable) when the desired input signal is an Autoregressive (AR) random process and when certain conditions are met. A prediction error method based adaptive feedback cancellation (PEM-AFC) is also proposed in [12] to identify the models for the desired signal and the feedback path simultaneously. However, in practice, many desired input signals, such as voiced speech and music, are not AR processes. Moreover, the conditions for identification may often be violated, for example, when insufficient filter length is used for modeling the desired input signal. In

Guilin Ma (gm@elektro.dtu.dk) is with Acoustic Technology, Department of Electrical Engineering, Building 352, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, and also with GN ReSound A/S, Algorithm Research, Lautrupbjerg 7, 2750 Ballerup, Denmark. Fredrik Gran (fgran@gnresound.dk) is with GN ReSound A/S, Algorithm Research, Lautrupbjerg 7, 2750 Ballerup, Denmark. Finn Jacobsen (fja@elektro.dtu.dk) is with Acoustic Technology, Department of Electrical Engineering, Building 352, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark. Finn T.Agerkvist (fa@elektro.dtu.dk) is with Acoustic Technology, Department of Electrical Engineering, Building 352, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.

these cases, bias remains in the estimate of the feedback path.

This paper proposes a new linear predictive coding based approach for reducing the bias. The idea is to generate a synthesis signal for the receiver input, which sounds perceptually similar to or possibly even the same as the original signal but is statistically uncorrelated with the desired input signal. It is shown that this approach reduces the bias significantly and improves the performance of the feedback cancellation system.

The paper is organized as follows: Section II describes the basic theory of linear predictive coding. In Section III, the band-limited linear predictive coding based adaptive feedback cancellation (BLPC-AFC) is proposed. An adaptive feedback cancellation system combining the BLPC-AFC and filtered-X adaptation is described in Section IV. In Section V, simulation results are presented and discussed. Concluding remarks are given in Section VI.

II. Linear predictive coding

Parametric representation of a spectrum by means of linear prediction (LP) is a powerful technique in speech and audio signal processing. Linear predictive coding (LPC) was developed for the purpose of speech compression in the 1960s [13]. After that, research on LPC vocoder resulted in the 2.4 kb/s secure-voice standard LPC10 [14]. However, the sound quality produced by LPC vocoder at low bit rates was not good enough for commercial telephony [13]. To provide high-quality speech at low bit rates, in the 1970s and 1980s, residual excited LPC (RELP), Multi-pulse LPC and code-excited LPC (CELP) were proposed to code the residual signal in better ways. The following subsections will briefly describe LPC for speech applications and its basis in the speech production model.

A. Discrete-time speech production model

The LPC-based vocoder, such as RELP, Multi-pulse LPC, CELP, etc., exploits the special properties of a classical discrete-time model of the speech production process, which is illustrated in Fig. 1. During unvoiced speech activity, the excitation source is flat-spectrum noise, modeled by a random noise generator; during periods of voiced speech activity, the excitation uses an estimate of the local pitch period to set an impulse train generator that drives a glottal pulse shaping filter. The excitation is later filtered by the vocal-tract filter and the lip radiation filter to produce the speech. This model, although not adequate for certain classes of phonemes such as voiced fricatives, has been successfully used in many speech analysis, coding and recognition tasks.

In general, modeling the transfer functions of vocal tract and lip radiation requires both zeros and poles. However, they can be well approximated by a complete all-pole model as illustrated in Fig. 2, which yields identical magnitude spectra to the true transfer function of the speech production process but might alter the phase characteristics. Applications have justified that correct spectral magnitude is frequently sufficient for coding, recognition and synthesis [15].

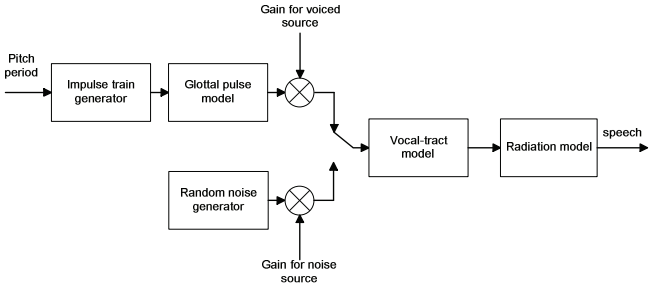


Fig. 1. A general discrete-time model for speech production [15].

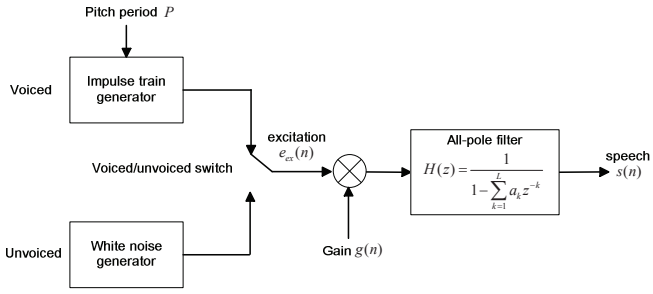


Fig. 2. An all-pole model for speech production [15]. The pitch period P, the type of excitation, the gain $g(n)$ and the all-pole filter $H(z)$ of order L are parameters to be estimated by linear prediction analysis. The excitation sequence is denoted by $e_{ex}(n)$, and $s(n)$ is the output speech from the production model.

In the all-pole speech model, the output speech $s(n)$ is generated with the excitation sequence $e_{ex}(n)$ in the following way:

$$s(n) = \mathbf{a}^T \mathbf{s}(n-1) + g(n) e_{ex}(n), \quad (1)$$

$$\mathbf{a} = [a_1, a_2, \cdots, a_L]^T, \quad (2)$$

$$\mathbf{s}(n) = [s(n), s(n-1), \cdots, s(n-L+1)]^T, \quad (3)$$

where $\mathbf{a}$ is the coefficient vector of the all-pole filter $H(z)$ of order L, and the superscript T denotes the transpose of a matrix.

Equations (1)-(3) suggest that except for the excitation term, $s(n)$ can be predicted using a linear combination of its past L values with the weights a_i's. The a_i, which characterize the all-pole filter, are usually estimated by an efficient computation technique called linear prediction analysis, which can be done in many ways, such as by using autocorrelation methods, covariance methods, etc., which will be described in Section II-C.

B. LPC vocoder

A typical diagram of LPC-based vocoder is given in Fig. 3. Speech in the transmission end $s(n)$ is first analyzed by LP analysis to estimate the set of coefficients of the all-pole filter, the pitch period, the gain parameter and the voiced/unvoiced parameter. These parameters are encoded for transmission through the communication channel. At the receiver end, the speech signal is synthesized in the way illustrated in Fig. 2 using the received parameters.

During the estimation of the parameters, the residual signal $s^p(n)$, also referred to as predicted error signal, can be ob-

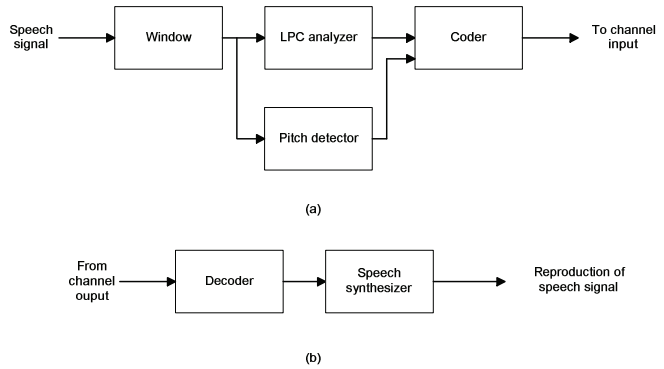


Fig. 3. Block diagram of a typical LPC vocoder: (a) transmitter; (b) receiver [16]. The window is typically 10-30 ms long. The parameters encoded for transmission over the communication channel: the set of coefficients computed by LPC analyzer, the pitch period, the gain parameter and the voiced/unvoiced parameter.

tained:

$$s^p(n) = s(n) - \hat{\mathbf{a}}^T \mathbf{s}(n-1), \quad (4)$$

$$\hat{\mathbf{a}} = [\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_L]^T, \quad (5)$$

$$\hat{H}(z) = \frac{1}{1 - \sum_{k=1}^{L} \hat{a}_k z^{-k}}, \quad (6)$$

where $\hat{H}(z)$ is the estimated all-pole filter of order L, the superscript p is used to denote the prediction error of the corresponding signal, and the denominator of $\hat{H}(z)$, which represents a Finite-Impulse-Impulse (FIR) filter, is also called the prediction error filter (PEF).

C. Linear prediction analysis

Linear prediction analysis is a way of estimating the AR model for a given signal. It is usually used in the LPC analyzer (see Fig. 3) to estimate the parameters, such as $\hat{\mathbf{a}}$. The LP analysis finds the set of coefficients of the all-pole filter by minimizing the mean-squared prediction error[1]:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} E\left\{|s^p(n)|^2\right\}, \quad (7)$$

where $E\{\cdot\}$ is the expectation operator, $s^p(n)$ is the prediction error/residual signal defined in (4).

Since the speech characteristics vary with time, the all-pole filter coefficients should be estimated by a short-term analysis, which minimizes the mean square of the prediction error over a segment of speech signal. The approaches for short-term LP analysis generally fall into two categories: the autocorrelation method and the covariance method. The autocorrelation method assumes that the samples outside the time segment are all zero. This assumption may result in a large prediction error at the beginning and end of the segment. To taper the segment and deemphasize that prediction error, a window (e.g., a Hamming window) is usually used. The covariance method, on the contrary, makes no assumptions about the values outside the segment and uses the true values.

For the autocorrelation method, the stability of the estimated all-pole filter can be guaranteed, whereas for the covariance method, it cannot be ensured. Therefore, the autocorrelation method is used in this paper. A well-known autocorrelation method is the Levinson-Durbin recursion algorithm [18] [19].

Another special type of methods for linear prediction is the lattice method. A typical lattice method is the Burg Lattice algorithm [20], which also yields stable all-pole filters.

[1]It should be noted that the minimization of the mean-squared prediction error $s^p(n)$ yields an all-pole system $\hat{H}(z)$ modeling the minimum-phase part of the true transfer function in Fig. 1 perfectly only during unvoiced signal segment. For voiced speech, although the model is not exact, the coefficients obtained still comprise a very useful and accurate representation of the speech signal [17].

III. Band-limited LPC vocoder for AFC

In this section, the bias problem associated with AFC is first explained through a steady-state analysis in Subsection III-A. Next, a new method based on a simplified LPC vocoder is proposed in Subsection III-B to reduce the bias. The developed LPC vocoder is band-limited to focus on bias reduction in the critical frequency region of the feedback path and to minimize the impact on sound quality. In the end, the steady-state analysis of the proposed BLPC-AFC is given in Subsection III-C.

A. Bias problem with AFC

A typical block diagram of AFC is illustrated in Fig. 4. The feedback path model $\hat{F}(z)$ usually consists of an adaptive FIR filter with the vector of coefficients $\hat{\mathbf{f}}$, i.e.,

$$\hat{F}(z) = \sum_{k=0}^{L_{\hat{F}}-1} \hat{f}_k z^{-k}, \quad (8)$$

$$\hat{\mathbf{f}} = \left[\hat{f}_0, \hat{f}_1, \cdots, \hat{f}_{L_{\hat{F}}-1}\right]^T, \quad (9)$$

where $L_{\hat{F}}$ is the length of the adaptive FIR filter $\hat{F}(z)$.

As pointed out in [4], the adaptation of this FIR filter to minimize the mean square of error $e(n)$ usually leads to a biased estimate when the desired input signal $x(n)$ is correlated with the receiver input signal $u(n)$. This can been shown from the steady-state analysis of the system, during which it is assumed that the feedback path is not varying and the input signal $x(n)$ is a wide-sense stationary process. Suppose that the feedback path $F(z)$ is also an FIR filter with coefficients vector $\mathbf{f}$ and is of the same order as the feedback path model $\hat{F}(z)$. The Wiener solution to the minimization of the mean-square error of the error signal $e(n)$ is:

$$\hat{\mathbf{f}} = \mathbf{R}_{\mathbf{uu}}^{-1} \mathbf{r}_{\mathbf{u}y} \quad (10)$$

$$= \mathbf{f} + \mathbf{R}_{\mathbf{uu}}^{-1} \mathbf{r}_{\mathbf{u}x}, \quad (11)$$

$$\mathbf{R}_{\mathbf{uu}} = E\left\{\mathbf{u}(n)\mathbf{u}^T(n)\right\}, \quad (12)$$

$$\mathbf{r}_{\mathbf{u}y} = E\left\{\mathbf{u}(n)y(n)\right\}, \quad (13)$$

$$\mathbf{u}(n) = \left[u(n), u(n-1), \cdots, u(n-L_{\hat{F}}+1)\right]^T, \quad (14)$$

where $\mathbf{r}_{\mathbf{u}x}$ and $\mathbf{r}_{\mathbf{u}f}$ are defined similarly as in (13). The term $\mathbf{R}_{\mathbf{uu}}^{-1}\mathbf{r}_{\mathbf{u}x}$ in (11) represents the bias of the estimate, which is related to the correlation between the desired input signal $x(n)$ and the processed hearing-aid signal $u(n)$. The magnitude of the bias depends strongly on the decaying speed of the autocorrelation function of $x(n)$, the forward-path delay and nonlinearity in the hearing-aid process $G(z)$. The bias problem is

particularly serious when the desired input signal $x(n)$ is tonal because the correlation function does not drop off.

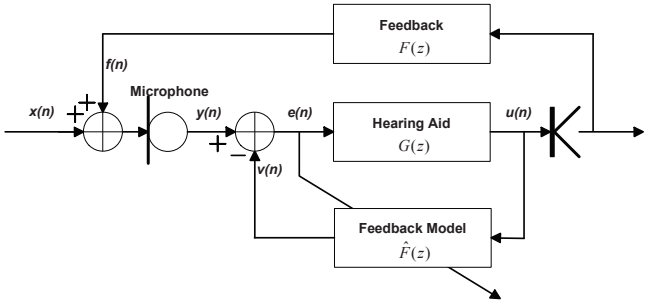


Fig. 4. General diagram of the adaptive feedback cancellation system. The input to the hearing-aid processing is $y(n)$, which is the sum of the desired input signal $x(n)$ and the feedback signal $f(n)$. The hearing-aid process is denoted as $G(z)$, and the processed hearing-aid signal is $u(n)$. The transfer function of the feedback path is $F(z)$, and $v(n)$ is the estimate of $f(n)$ generated by the modeled feedback path $\hat{F}(z)$.

B. Band-limited LPC vocoder for AFC

To reduce the bias, several approaches have been proposed as mentioned in Section I. Here, a new method to decorrelate $x(n)$ and $u(n)$ using a band-limited LPC is proposed.

The main idea is to create a synthetic replica of the processed hearing-aid signal $u(n)$, which is statistically uncorrelated with $x(n)$ but still sounds perceptually close or identical to $u(n)$. To achieve this, a simplified LPC vocoder is adopted, which consists of three steps: First, LP analysis is performed on $u(n)$ to estimate the all-pole model for $u(n)$; then the residual signal is replaced with a white noise sequence of the same variance as the residual signal of $u(n)$; in the end, the noise sequence drives the obtained all-pole system to synthesize a new signal $u_{syn}(n)$ for the receiver to output, which maintains the magnitude spectrum of $u(n)$ but is uncorrelated with $x(n)$.

Compared with a standard LPC vocoder, such as LPC10, this simplified vocoder has a great advantage in terms of computation load since it does not need any voiced activity detection and pitch estimation. It also removes the long-term bias in the adaptation completely since $u_{syn}(n)$ is uncorrelated with $x(n)$ (cf. Equation (11)). However, as mentioned in Section II, voiced speech is synthesized with an impulse train[2]. With only white noise driven, the synthesis of voice speech will degrade sound quality significantly.

To reduce this effect, a band-limited LPC vocoder (BLPC) is proposed based on the characteristics of the feedback path and the performance of AFC in practice. Previous research has shown that the magnitude of the frequency response of the feedback path is usually much higher in the region above 2 kHz than that below 2 kHz [21] (cf. Fig. 7 in Section V). For most hearing-aid users, the prescribed forward-path gain is also higher at high frequencies than at low frequencies. Therefore in practice, the AFC fails to prevent whistling at high frequencies in most cases. Moreover, since the feedback is usually very weak at low frequencies, special methods can be used in AFC to prevent whistling resulted by the bias at low frequencies. For example, high-pass filters can be used in front of the adaptation of the feedback model to reduce the effect of the bias at low frequencies [22]. Thus the bias problem is prominent mainly at high frequencies, and the reduction of bias, as a means to improve the performance of AFC, is mainly needed in the region above 2 kHz.

To decorrelate $x(n)$ and $u(n)$ at high frequencies, the synthesized signal $u_{syn}(n)$ is only needed in the high-frequency region while the low-frequency part of the original signal $u(n)$ can be maintained without any modification. This consideration results in a band-limited LPC vocoder based AFC (BLPC-AFC) as illustrated in Fig. 5. The processed hearing-aid signal $u(n)$ is input to the LP analysis to estimate the all-pole filter $\hat{H}(z)$ and the residual gain $g(n)$ using one of the methods that yield stable models described in Section II-C. The residual gain $g(n)$ approximates the standard deviation of the prediction error/residual signal so that the power of the original signal is maintained. The way of estimating $g(n)$ will be given in Section V. In the LP synthesis stage, a unit-variance white noise excitation $w(n)$ is used to drive the estimated all-pole filter $\hat{H}(z)$ with an amplification of $g(n)$ to produce the synthesize signal $u_{syn}(n)$, which is high-pass filtered afterwards to obtain the high-frequency component $u_{syn\_hp}(n)$. In the end, $u_{syn\_hp}(n)$ is added to $u_{lp}(n)$, the low-pass filtered $u(n)$, to obtain a new signal $q(n)$ for the receiver to output, i.e.,

$$q(n) = u_{lp}(n) + u_{syn\_hp}(n). \tag{15}$$

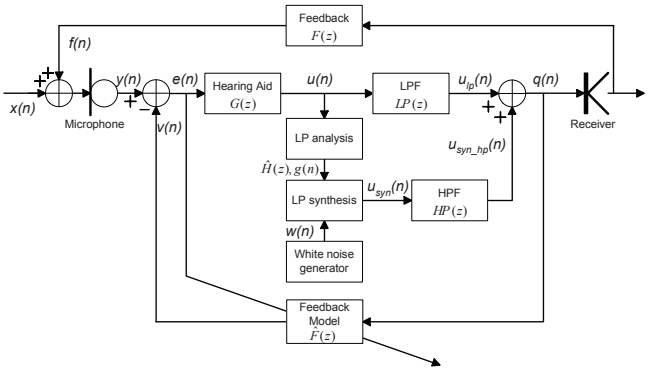


Fig. 5. Diagram of adaptive feedback cancellation with band-limited LPC vocoder. LPF is the low-pass filter with the transfer function LP(z), and HPF is the high-pass filter with the transfer function HP(z).

By keeping the low-frequency signal intact, the sound quality is improved significantly at least for speech signal as most energy of the speech signal is concentrated at low frequencies. The BLPC vocoder proposed here actually resembles the RELP vocoder, in which the residual signal below 1 kHz is transmitted as the excitation sequence for the LP synthesis. The differences between the RELP and the BLPC vocoder lie in two aspects: First, RELP typically has a cut-off frequency at 1 kHz while BLPC has a cut-off frequency at 2 kHz, which means the sound quality of BLPC below 2 kHz is better than that in RELP; Secondly, in RELP, the high-frequency signal is restored in some nonlinear manner, typically with a rectifier [15], whereas, BLPC restores it with white noise excitation. This implies that RELP may still recover the formants above 1 kHz to

[2]Strictly speaking, a phase altered version of an impulse train.

some extent while BLPC loses the fine structure of the formants above 2 kHz. However, by keeping the original signal intact below 2 kHz, BLPC has already maintained the first formants and most of the second formants of vowels.

C. Steady-state analysis of BLPC-AFC

The bias of BLPC-AFC can be calculated from a steady-state analysis of the system by assuming that a Least-Square solution is obtained (cf. Equation (11)):

$$\begin{aligned} \text{Bias} &= \mathbf{R}_{\mathbf{qq}}^{-1}\mathbf{r}_{\mathbf{q}x}, && (16)\\ &= \mathbf{R}_{\mathbf{qq}}^{-1}E\left\{\left[\mathbf{u}_{lp}(n)+\mathbf{u}_{syn\_hp}(n)\right]x(n)\right\}, && (17)\\ &= \mathbf{R}_{\mathbf{qq}}^{-1}E\left\{\mathbf{u}_{lp}(n)x(n)\right\}, && (18)\\ \mathbf{q}(n) &= \left[q(n),q(n-1),\cdots,q(n-L_{\hat{F}}+1)\right]^T, && (19)\end{aligned}$$

where the vectors $\mathbf{u}_{lp}(n)$ and $\mathbf{u}_{syn\_hp}(n)$ are defined similarly as in (19). From (17) to (18), the fact is utilized that the synthesized signal $u_{syn\_hp}(n)$ generated from a white noise sequence is statistically uncorrelated with the desired input signal $x(n)$.

Equation (18) shows that the high-frequency bias is removed. Although the bias remains at low frequencies, it usually does not result in any problem because the feedback cancellation system, in most cases, handles the low-frequency bias very well but fails to prevent whistling at high frequencies as mentioned in Section III-B.

IV. Band-limited LPC vocoder for AFC with filtered-X adaptation

The proposed BLPC-AFC can be further combined with two adaptive decorrelation filters (ADF) in the feedback cancellation path to reduce the short-time correlation in the high-frequency region and yield more accurate estimate of the feedback and is therefore called "BLPC-FxAFC" algorithm.

A. The use of the filtered-X adaptation in BLPC-AFC

The BLPC vocoder helps to remove the long-term bias in the high-frequency region as shown in Section III-C. However, short-term correlation still exists especially for tonal signals, which may lead the system adaptation in a wrong direction when the adaptation algorithm, such as Normalized-Lease-Mean-Square (NLMS), uses data within a short observation window. To reduce this short-term correlation, two decorrelation filters can be introduced in the feedback cancellation path.

Suppose the estimated all-pole filter $\hat{H}(z)$ is obtained in the LP analysis stage. The inverse of $\hat{H}(z)$ is an FIR filter, which is also referred to as the prediction error filter (PEF) as mentioned in Section II-B. Denote this PEF as $\hat{A}(z)$, then

$$\hat{A}(z) = 1 - \sum_{k=1}^{L} \hat{a}_k z^{-k}. \qquad (20)$$

The adaptation of the feedback path model is based on the receiver input signal $q(n)$ and the error signal $e(n)$. If both signals are filtered with a decorrelation filter $\hat{A}(z)$ before entering the adaptation, then a structure identical to filtered-X adaptation[3] is achieved [23,24]. The advantage of using $\hat{A}(z)$ to filter $q(n)$ and $e(n)$ is that at the receiver end, the high-frequency component of the filtered signal of $q(n)$ will be exactly the high-pass filtered white noise sequence that is used to generate the synthesized signal $u_{syn\_hp}(n)$, i.e. $w(n)$ filtered by $HP(z)$, if the ADFs and the estimated $\hat{A}(z)$ are synchronized perfectly. The temporal correlation between $x(n)$ and $q(n)$ at high frequencies can be decreased significantly in this way, which will be shown by an example in Section V. Since $\hat{H}(z)$ is estimated on the broadband signal $u(n)$, the inverse filter $\hat{A}(z)$ used in the filtered-X will whiten the two signals $x(n)$ and $q(n)$ at low frequencies to some extent[4] and help to reduce the temporal correlation. The filtered-X adaptation based BLPC-AFC, BLPC-FxAFC, is illustrated in Fig. 6, where the PEF $\hat{A}(z)$, estimated from LP analysis of the processed hearing-aid signal $u(n)$, is copied to the two ADFs to generate the prediction errors $e^p(n)$ and $q^p(n)$ for adaptation in the feedback model. Since the two ADFs use the same filter $\hat{A}(z)$, the phase misalignment between these two filters is zero and therefore the requirement of phase misalignment for stable adaptation [21] of the filtered-X algorithm is always satisfied. However, due to the group delay associated with the ADFs, the filtered-X algorithm may become unstable if the coefficients of the estimated feedback path $\hat{F}(z)$ change too fast [12].

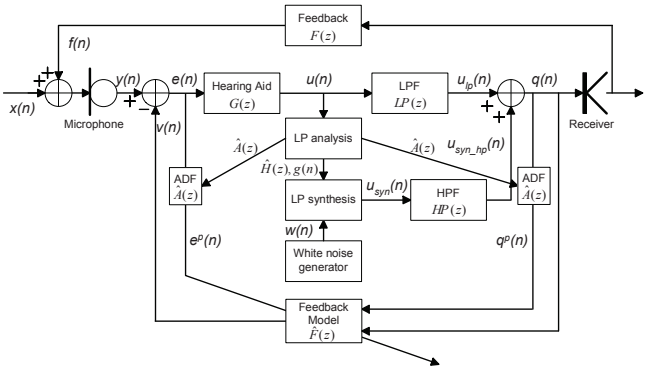


Fig. 6. Diagram of the feedback cancellation system with band-limited LPC vocoder and filtered-X adaptation. The receiver input $q(n)$ and the prediction error signal $q^p(n)$ are both input to the feedback model. The former is used to generate the feedback estimation signal $v(n)$ and the latter is used to update the feedback model together with $e^p(n)$.

B. Steady-state analysis

In the proposed BLPC-FxAFC, the estimated feedback path in the steady state, assuming that the Least-Square solution has

[3]It should be noted that $\hat{A}(z)$ is dependent on the characteristics of the incoming signal. Therefore the two decorrelation filters, which use the coefficients of $\hat{A}(z)$, are actually adaptive.

[4]The low-frequency whitening will not be as effective as that at high frequencies unless the desired input signal $x(n)$ is an AR random process.

been obtained, is as follows,

$$\begin{aligned}
\hat{\mathbf{f}} &= \mathbf{R}^{-1}_{\mathbf{q}^p\mathbf{q}^p}\mathbf{r}_{\mathbf{q}^p y^p} && (21)\\
&= \mathbf{R}^{-1}_{\mathbf{q}^p\mathbf{q}^p}\mathbf{r}_{\mathbf{q}^p f^p} + \mathbf{R}^{-1}_{\mathbf{q}^p\mathbf{q}^p}\mathbf{r}_{\mathbf{q}^p x^p}, && (22)\\
\mathbf{q}^p(n) &= \left[q^p(n), q^p(n-1), \cdots, q^p(n-L_{\hat{F}}+1)\right]^T && (23)\\
q^p(n) &= q(n) - \hat{\mathbf{a}}^T\mathbf{q}_L(n-1), && (24)\\
\mathbf{q}_L(n) &= \left[q(n), q(n-1), \cdots, q(n-L+1)\right], && (25)
\end{aligned}$$

where $\mathbf{R}_{\mathbf{q}^p\mathbf{q}^p}$ is defined similarly as in (12), $\mathbf{r}_{\mathbf{q}^p y^p}$, $\mathbf{r}_{\mathbf{q}^p f^p}$ and $\mathbf{r}_{\mathbf{q}^p x^p}$ are defined similarly as in (13), the superscript p denotes the prediction error of the corresponding signal and the prediction errors $y^p(n)$, $x^p(n)$ and $f^p(n)$ are defined similarly as in (24).

In (22), the first term is essentially the steady-state optimal solution to a filtered-X Wiener filtering, which approximates the true feedback path as long as the filter length of $\hat{F}(z)$ is sufficiently large for the feedback path modeling. In the second term, the source of bias $\mathbf{r}_{\mathbf{q}^p x^p}$ and can be further expanded:

$$\begin{aligned}
\mathbf{r}_{\mathbf{q}^p x^p} =& E\left\{\left[\mathbf{u}^p_{syn_hp}(n) + \mathbf{u}^p_{lp}(n)\right]x^p(n)\right\}, && (26)\\
=& E\left\{\mathbf{u}^p_{lp}(n)x^p(n)\right\}, && (27)
\end{aligned}$$

where $\mathbf{u}^p_{lp}(n)$ and $\mathbf{u}^p_{syn\_hp}(n)$ are defined similarly as in (23). From (26) to (27), the fact is utilized that the synthesized signal generated with a white noise sequence is uncorrelated with $x(n)$ and $x^p(n)$.

Equations (26)-(27) show that at high frequencies, the bias can be eliminated as long as the filter length of $\hat{F}(z)$ is sufficient. The feedback estimate at high frequencies is not influenced by the estimation of $\hat{A}(z)$ even when $\hat{A}(z)$ has an insufficient order for modeling $u(n)$ or when it does not model $u(n)$ accurately.[5]

It should be noted that the steady-state analysis of BLPC-AFC in Section III-C has shown similar results, i.e., the elimination of the bias at high frequencies. Therefore the advantage of BLPC-FxAFC is not expected in the least-square solution using long-term steady-state data but expected in the practical situation where the adaptation of the feedback model uses data within a short observation window. This will be further explained in Section V-A.

C. Comparison of BLPC-AFC, BLPC-FxAFC and PEM-AFC

Both BLPC-AFC and BLPC-FxAFC can eliminate the long-term bias in the high-frequency region as long as the filter $\hat{F}(z)$ is long enough to model the feedback path $F(z)$. BLPC-FxAFC can further reduce the short-term correlation especially at high frequencies since the prediction error is a high-pass filtered white noise sequence at the receiver end. This can yield a better estimate of the feedback path.

The filtered-X algorithm used in the proposed BLPC-FxAFC is similar to the PEM-AFC proposed in [12] to some extent because both use linear prediction coefficients to decorrelate the error signal and the receiver input signal. The difference lies in the fact that PEM-AFC minimizes the prediction error of the input signal of the hearing-aid process $e(n)$, whereas BLPC-FxAFC minimizes the prediction error of the output signal of the hearing-aid process $u(n)$. If it is assumed that the forward-path hearing aid process $G(z)$ contains only a delay d_G and a constant linear amplification g_G, the position of linear prediction after the hearing-aid process does not result in any difference in the steady-state performance [25]. In this sense, the proposed BLPC-FxAFC can also be roughly interpreted as the combination of a BLPC-AFC with reduced short-term correlation in the high-frequency region and a modified PEM-AFC in the low-frequency region.

PEM-AFC removes the bias only when the desired input signal $x(n)$ is an AR random process and when certain conditions are met [12]. For a large set of real-life signals, such as voiced speech and tonal music, which can hardly be modeled by an all-pole filter, PEM-AFC still suffers from a biased solution because the prediction error signals $u^p(n)$ and $e^p(n)$ are not white [12]. Moreover, under-modeling of the desired input signal $x(n)$ may also introduce bias into the estimation. For these two cases, BLPC-FxAFC can still be useful in removing the bias in the high-frequency region where feedback usually occurs.

V. Simulations and Discussion

To evaluate and compare the performance of the algorithms, simulations are carried out for AFC, BLPC-AFC, PEM-AFC, Filtered-X AFC (FxAFC) and BLPC-FxAFC. The FxAFC uses the same filtered-X approach as used in BLPC-FxAFC but does not involve the synthesis stage. It can also be regarded as a modified PEM-AFC with linear prediction placed at the receiver end.

The five methods are simulated with a sampling frequency of 16 kHz. The processing is block based with a block size of 24 samples, corresponding to 1.5 ms. The forward path $G(z)$ consists of a delay d_G of 24 samples and an adjustable linear gain g_G. Most hearing impaired people have greater hearing loss at high frequencies. Therefore the prescribed gain in the forward path will also be higher at high frequencies. This gain setting has become one of the biggest challenges for feedback cancellation in practice. To simulate a realistic gain setting in the hearing aids and also to test the performance of the algorithms with high gains at frequencies where feedback oscillation usually occurs, the forward-path gain g_G is set to 15 dB at frequencies below 2 kHz and 35 dB above 2 kHz in all the simulations.

In the simulations, the feedback path is an FIR filter with 50 orders obtained from the measurement of a commercial behind-the-ear (BTE) hearing aid, ReSound Metrix MX70-DVI. The frequency response of the feedback path is illustrated in Fig. 7, which has large magnitude responses from 2 kHz to 7 kHz. The maximum stable gain without feedback canceller is around 15 dB at 3.3 kHz. The feedback model $\hat{F}(z)$ consists of an adaptive FIR filter of 50 orders, which is initialized as the true feedback path to show how the estimate of the feedback path drifts away from the true feedback path due to the bias problem. This initialization of the filter is also considered as a result of a common fitting procedure for the feedback cancellation in the

[5] The under-modeling or wrong modeling does not introduce any bias but will degrade sound quality of the synthesized signal.

industry [3], in which the true feedback path is measured and used as the starting point and/or constraint of the adaptation.

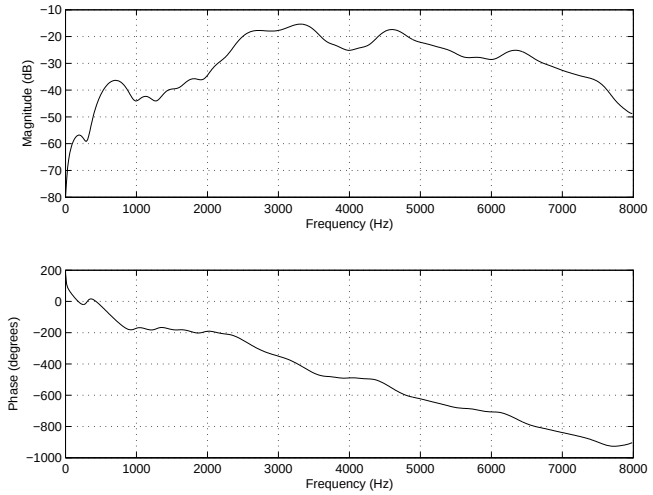


Fig. 7. The frequency response of the feedback path of 50 orders based on the measurement of a commercial BTE hearing aid: ReSound Metrix.

The adaptive filter is updated by a block-based NLMS algorithm, which is a modified block LMS algorithm [16]. In AFC and BLPC-AFC, the update is performed as follows,

$$\hat{\mathbf{f}}(m+1) = \hat{\mathbf{f}}(m) + \alpha(m) \sum_{i=0}^{B-1} \mathbf{q}(mB+i)e(mB+i), \quad (28)$$

$$\alpha(m) = \frac{\mu}{\delta + \sum_{i=0}^{i=B-1} |q(mB+i)|^2}, \quad (29)$$

where $\mathbf{q}(mB+i)$ is defined in (19), m is the block index, B is the block size and equals 24 samples, μ is the step-size parameter and set to 0.002 in the simulations, and δ is set to a small value to overcome numerical difficulties. For PEM-AFC, FxAFC and BLPC-FxAFC, the update is similar

$$\hat{\mathbf{f}}(m+1) = \hat{\mathbf{f}}(m) + \alpha(m) \sum_{i=0}^{B-1} \mathbf{q}^p(mB+i)e^p(mB+i), \quad (30)$$

$$\alpha(m) = \frac{\mu}{\delta + \sum_{i=0}^{i=B-1} |q^p(mB+i)|^2}, \quad (31)$$

where $\mathbf{q}^p(mB+i)$ is defined in (23), $q^p(mB+i)$ is defined in (24).

In PEM-AFC, FxAFC and BLPC-FxAFC, the PEF $\hat{A}(z)$ are of the length 21, which is the same as used in [12]. The autocorrelation method Levinson-Durbin algorithm, which yields stable models, is used with an analysis window length of 10.5 ms, corresponding to 168 samples or 7 blocks. The $\hat{H}(z)$ is updated for every new block. Therefore the linear prediction for the current block is based on the data in the current block and in the six previous blocks. The residual gain for block m is estimated in the following way:

$$g(m) = \sqrt{\frac{1}{B} \sum_{i=0}^{B-1} |u^p(mB+i)|^2}, \quad (32)$$

where $u^p(mB+i)$ is defined similarly as in (24), $\hat{\mathbf{a}}(m)$ is the estimated coefficients of the all-pole model at block m. The residual gain $g(m)$ makes sure that the power of the residual signal $u^p(n)$ in each block is the same as the variance of the noise sequence used for synthesis, which is done in the way as illustrated in Fig. 2:

$$u_{syn}(mB+i) = \hat{\mathbf{a}}(m)^T \mathbf{u}_{syn,L}(mB+i-1) + g(m)w(mB+i), \quad (33)$$

where $i = 0, 2, \cdots, B-1$ and $\mathbf{u}_{syn,L}(mB+i-1)$ is defined similarly as in (25).

The high-pass filter $HP(z)$ is a 40-order FIR filter and has a cut-off frequency of 2 kHz. It is designed with the classical windowed linear-phase FIR digital filter design method [26] using a hamming window. The low-pass filter $LP(z)$ is also 40-order and is the strict complementary filter of $HP(z)$, i.e.,

$$LP(z) = z^{-n_0} - HP(z), \quad (34)$$

where n_0 is the group delay of the designed $HP(z)$ and equals 20 samples. The additional delay introduced by n_0 in the forward path of BLPC-AFC and BLPC-FxAFC is accordingly added in the forward path of AFC, PEM-AFC and FxAFC so that the performance comparison between these algorithms is not influenced by the overall forward-path delay. The overall forward-path delay therefore is the sum of d_G and n_0 for all the algorithms. The frequency responses of the low- and high-pass filters are shown in Fig. 8.

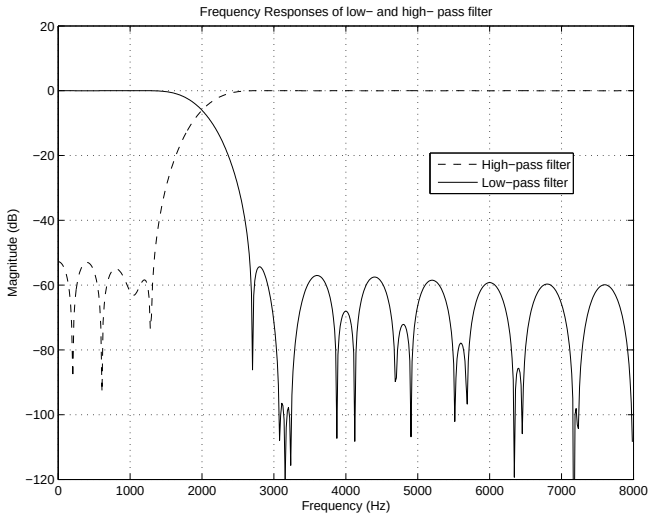


Fig. 8. The frequency responses of the complimentary low- and high-pass filters.

The performance of the algorithms is evaluated by the misalignment between the true feedback path $F(z)$ and the modeled feedback path $\hat{F}(z)$. The misalignment is calculated at frequencies above 2 kHz to quantify the modeling error in the critical frequency region where feedback oscillation usually occurs and to show the effects of BLPC vocoder. The misalignment above 2 kHz is denoted as ς, which is computed in the

frequency domain as

$$\varsigma = \frac{\sum_{k=N_0}^{N_f-1} \left| F(e^{j2\pi(k/N_f)}) - \hat{F}(e^{j2\pi(k/N_f)}) \right|^2}{\sum_{k=N_0}^{N_f-1} \left| F(e^{j2\pi(k/N_f)}) \right|^2}, \quad (35)$$

$$N_0 = \lceil 2000/(fs/Nf) \rceil \quad (36)$$

where $\lceil \cdot \rceil$ is the ceiling function to get the smallest integer not less than the value in the brackets, N_f equals the number of frequency points, which is 1024 in this paper, and fs is sampling frequency in Hz, which equals 16000 in the simulations. Therefore, N_0 is calculated as 128.

A. Simulation results with a stationary AR signal input

To examine the performance of the algorithms, a stationary AR random process of 8 seconds, which is generated by a 20-order all-pole filter $H_1(z)$, also called "signal model", is used as the input signal in the first test case. The power spectrum density (PSD) of the AR signal, which is shown in Fig. 9, exhibits sharp peaks. Therefore the bias problem is expected to be serious for the conventional AFC. The misalignment above 2 kHz is depicted in Fig. 10.

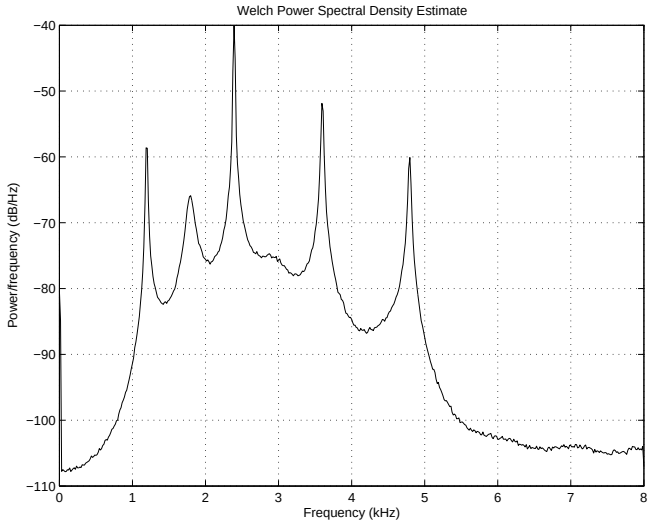


Fig. 9. The power spectral density of the 20-order AR random process.

As can be seen from Fig. 10, due to the bias problem, AFC exhibits the largest misalignment. BLPC-AFC lowers the misalignment by around 7 dB on average. However, its misalignment has the largest fluctuations because the short-term correlation between the synthesized AR signal and the original AR signal has a large variance.

To illustrate this short-term correlation, suppose there are two AR signals $r_1(n)$ and $r_2(n)$ generated from the same signal model as that of the test signal, i.e. $H_1(z)$, but with two different white noise sequences $n_1(n)$ and $n_2(n)$ respectively. The two signals are both of the length 1000 samples. The normalized cross-correlation[6] of $r_1(n)$ and $r_2(n)$, auto-correlation of $r_1(n)$ and cross-correlation of $n_1(n)$ and $n_2(n)$ are illustrated in Fig 11(a). As shown in the figure, the decaying speed of the auto-correlation of $r_1(n)$ is very slow and therefore the short delay in the hearing-aid forward path is not sufficient to reduce the correlation between $r_1(n)$ and its delayed replica. When the delayed replica of $r_1(n)$ is replaced by another uncorrelated AR process $r_2(n)$, the short-term correlation gets smaller but is still high. However, the cross-correlation between $n_1(n)$ and $n_2(n)$ is much smaller. This explains why BLPC-FxAFC yields much better performance than BLPC-AFC in Fig. 10. The temporal correlation between $r_1(n)$ and $r_2(n)$ exhibits a very large variance as shown in Fig 11(b), where the cross-correlation between $r_1(n)$ and 50 realizations of $r_2(n)$ and the averaged cross-correlation are illustrated. The 50 realizations of $r_2(n)$ are obtained by using 50 different white noise sequences. The large variance of the temporal correlation between $r_1(n)$ and $r_2(n)$ results in large fluctuations in the misalignment curve of BLPC-AFC. It can also be seen that the average cross-correlation is much smaller, which implies that the long-term bias can be removed by BLPC-AFC.

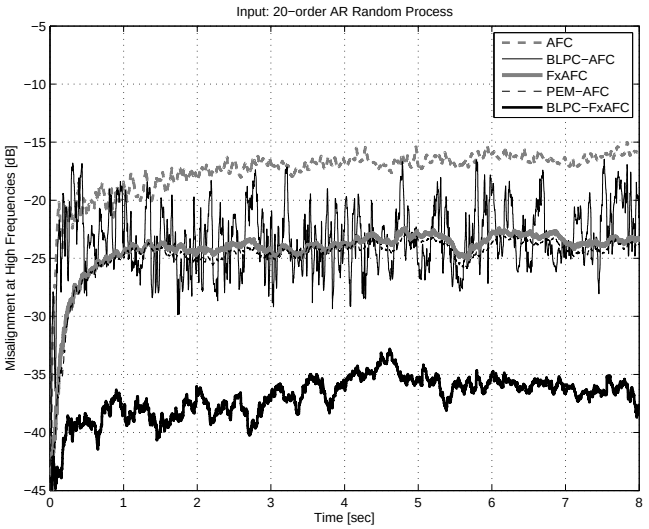


Fig. 10. The misalignment at high frequencies when a 20-order AR random process is used as the desired input signal.

Figure 10 also shows that the performance of FxAFC is very close to that of PEM-AFC, and the performance of BLPC-FxAFC is much better than that of FxAFC and PEM-AFC. This is because the on-line estimation of the signal model from a short observation window exhibits variation, which will result in non-white prediction error and short-term bias, and therefore limits the performance of FxAFC and PEM-AFC. For BLPC-FxAFC, although this problem also exists, the performance is not influenced too much because the prediction error at the receiver end is always a high-pass filtered white noise sequence as pointed out in Section IV-A.

[6]The normalized cross-correlation refers to the cross-correlation between two normalized sequences. Each sequence is normalized so its autocorrelation at zero lag is unity.

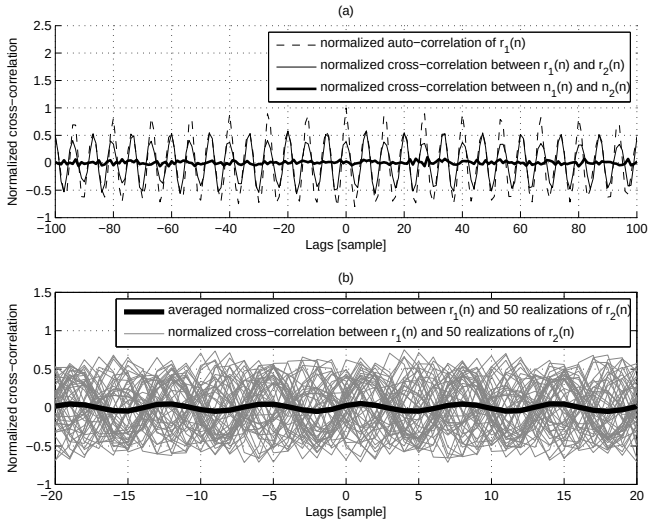


Fig. 11. (a) The normalized cross-correlation between two 20-order AR random processes $r_1(n)$ and $r_2(n)$, between their corresponding white noise sequences $n_1(n)$ and $n_2(n)$, and the auto-correlation of $r_1(n)$. (b) The normalized cross-correlation between $r_1(n)$ and 50 realizations of $r_2(n)$ and the averaged normalized cross-correlation.

B. Simulation results with a speech signal input

In the second test case, an 8-second sample of female speech is used as the input signal. The speech signal and the misalignment above 2 kHz is illustrated in Fig. 12.

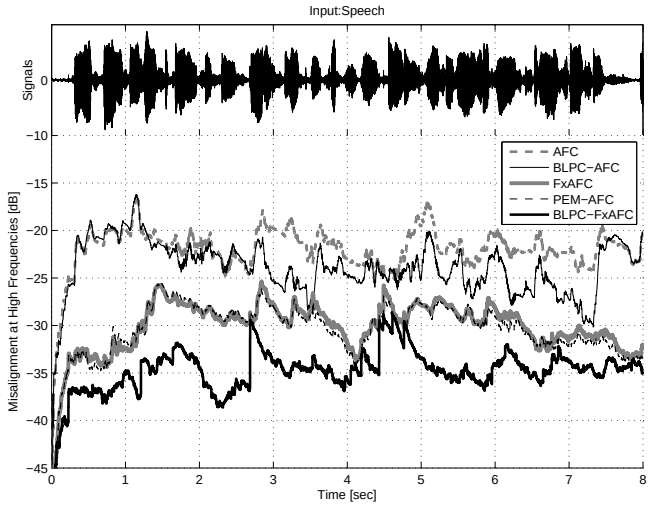


Fig. 12. The speech signal and the misalignments at high frequencies.

Figure 12 shows that the misalignment of BLPC-AFC is around 2-3 dB lower than that of AFC. The performance of FxAFC is again very close to that of PEM-AFC. Compared with FxAFC and PEM-AFC, BLPC-FxAFC reduces the misalignment by around 5 dB on average. This shows that BLPC vocoder helps to improve the estimation accuracy of the feedback path. The filtered-X based algorithms yield better performance as expected.

The difference in the performance between the best and the worst algorithms is smaller than that in the previous test case. This is because the speech signal is generally not very correlated with itself. Although during the periods of voiced speech the autocorrelation is significant, the voiced state usually does not last very long, and thus the buildup of bias is not large if the step-size parameter of the adaptation in the feedback cancellation is small enough and/or when a sufficient delay is introduced in the hearing-aid process. Therefore the bias problem tends to be smaller with speech input signal.

It can also be noticed that all the curves exhibit significant fluctuations. This is due to the dynamic nature of speech. The speech signal is only stationary for 10-20 ms and switches frequently between voiced state, unvoiced state and pauses.

The misalignment of FxAFC, PEM-AFC and BLPC-FxAFC fluctuates more than that of AFC and BLPC-AFC. This actually happens in the transient part of speech, during which the analysis frame of linear prediction contains a segment of non-stationary signal. Linear prediction with non-stationary data will result in an inaccurate model. For FxAFC and PEM-AFC, using the inverse of this inaccurate model as the ADFs does not whiten $e(n)$ at the microphone side and $q(n)$ at the receiver input, and may even color the signal and introduce short-term bias in the adaptation. For BLPC-FxAFC, this inaccurate modeling also occurs, but the misalignment is smaller than FxAFC and PEM-AFC because at the receiver end the signal after the decorrelation filter $q^p(n)$ is white at high frequencies.

C. Simulation results with a music signal input

In the third case, an 8-second sample of flute music is used as the input signal. The spectrogram of the music signal is illustrated in Fig. 13, which shows that the music signal is very tonal and therefore very challenging for feedback cancellation systems. The spectrogram is normalized so that the maximum magnitude is 0 dB.

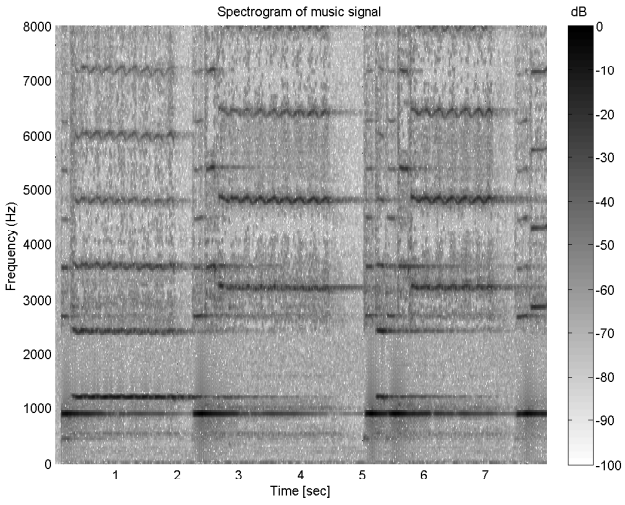


Fig. 13. The spectrogram of the 8-second flute music signal which is normalized so that the maximum peak is 0 dB.

The misalignment above 2 kHz is shown in Fig. 14. BLPC-AFC and AFC both yield large misalignment although BLPC-AFC is slightly better. This is because the short-time correlation for the tonal flute music input is very high even when the original signal is replaced by the synthesized signal generated with a white noise sequence (cf. Fig. 11). It takes a long time to

average out this high temporal correlation with an NLMS adaptation algorithm. In fact, feedback whistling happens for AFC and BLPC-AFC at some places of the output signal.

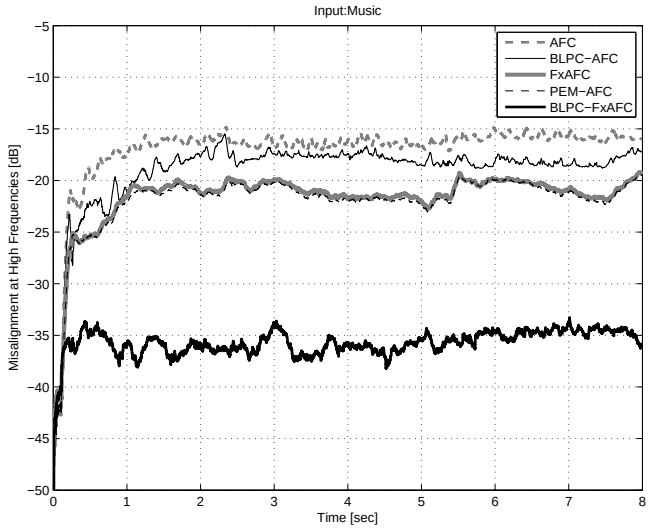


Fig. 14. The misalignment at high frequencies when the flute music is used as the desired input signal.

The performance of FxAFC and PEM-AFC is very similar. Thanks to the two ADFs, they both give a better performance than AFC and BLPC-AFC. But the remaining bias still exists because the flute signal is not a perfect AR process. BLPC-FxAFC shows a significant improvement in the performance over the other three methods because of both the replacement with an uncorrelated signal and the filtered-X adaptation.

D. Remarks on sound quality

The sound quality of the synthetic signals using the BLPC with the same parameters and linear prediction algorithm (Levinson-Durbin) in the simulation has been evaluated subjectively by the authors.

For speech samples, the overall sound quality is degraded very little although the difference between the original speech and the synthesized speech can still be perceived. For hearing impaired listeners, it is very likely that even this difference can hardly be detected. During the transient part of speech, noticeable effects due to the inaccurate modeling as mentioned in Section V-B are very rare. This is mainly because of the characteristics and the parameters of BLPC. Firstly, a relatively short analysis window ($\sim$10ms) with heavy overlapping (85.7%) is used in the BLPC to get a good time resolution, which is one of the easiest ways to reduce transient effect [27]. Secondly, the BLPC synthesis only takes effect at high frequencies and therefore the dominant energy of speech, which is usually located at low frequencies, may mask the error signal resulted from the inaccurate modeling in the high-frequency region. Lastly, since the synthesis is driven by a white noise sequence instead of an impulse train, only noise could be heard when inaccurate modeling happens instead of other unpleasant artifacts. When the microphone noise and ambient noise are present, this noise due to inaccurate modeling sounds even weaker or inaudible.

For tonal music samples, the degradation of sound quality depends on the characteristics of the signals. For signals with a few sharp peaks spaced sparsely in the high-frequency spectrogram, such as the flute music sample, although the sound quality is not preserved as well as for speech, it is not degraded very much due to a high-order all-pole filter used to model the spectrogram. For signals with a lot of peaks at high frequencies of the spectrogram or very complicated high-frequency spectrogram, the sound quality is degraded to some extent because the modeling fails to capture the envelope of the spectrogram.

The thorough evaluation of the sound quality is not the scope of this paper, which aims at technical description of the algorithms and performance evaluation. The perceptual validity of these preliminary findings is best addressed using a clinical trial or/and an objective measure which will be the subject of future studies.

VI. Conclusion and future directions

In this paper, a new approach to the bias problem encountered in adaptive feedback cancellation in hearing aids is presented. The main idea of the method is to replace the receiver input signal with a synthesized signal, which sounds perceptually similar to or even identical to the original signal but is statistically uncorrelated with the desired input signal.

To achieve this, a BLPC vocoder is proposed, which is based on band-limited linear predictive coding of the processed hearing-aid signal. To obtain effective decorrelation, impulse trains are not used for excitation as in conventional LPC-based vocoders during voiced speech. Instead, a white noise sequence is always used to drive the estimated signal model to generate the synthesized signal. Based on the facts that the magnitude of the frequency response of the feedback path is usually much higher in the high-frequency region and that the AFC usually breaks down at high frequencies, the signal replacement is performed at high frequencies to focus on the critical frequency region to improve the performance of AFC and also to reduce the degradation in sound quality.

The BLPC vocoder can be used on top of a conventional AFC to yield BLPC-AFC to reduce the long-term bias. Moreover, the BLPC-AFC method can be further combined with filtered-X adaptation to get BLPC-FxAFC, which can effectively reduce the short-term bias. The proposed BLPC-FxAFC can also be regarded as a modified version of the previously proposed PEM-AFC approach combined with BLPC vocoder.

The simulation results show that BLPC is effective in reducing the bias and the misalignment between the estimated and the real feedback paths. The BLPC-FxAFC method has the best performance for all the test signals.

The BLPC vocoder has a cutoff frequency at 2 kHz, which avoids severe degradation of sound quality. According to the subjective evaluation of the authors, the sound quality is very well preserved for speech. For many music signals with only a few peaks sparsely spaced at high-frequency spectrogram, the sound quality is not degraded very much either. A clinical trial and/or objective measure is still needed in the future to verify these findings, which will be the subject of future research. Also, it is found that the dynamic nature of speech makes it

hard for PEF to keep up and to effectively decorrelate the signals. Two possible approaches could be investigated to improve the dynamic AR modeling in the future: The first approach is to use other time-varying LPC techniques, such as the methods proposed in [28]; the second approach is to use a detector of speech transition to adjust the position and length of the analysis window of linear prediction.

REFERENCES

[1] B. Rafaely, M. Roccasalva-Firenze, and E. Payne, "Feedback path variability modeling for robust hearing aids," *J. Acoust. Soc. Am.*, vol. 107(5), pp. 2665–2673, 2000.

[2] S. F. Lybarger, "Acoustic feedback control," *The Vanderbilt Hearing-Aid Report, Studebaker and Bess, Eds. Upper Darby, PA: Monographs in Contemporary Audiology*, pp. 87–90, 1982.

[3] J. M. Kates, "Constrained adaptation for feedback cancellation in hearing aids," *Journal of the Acoustical Society of America*, vol. 106(2), pp. 1010–1019, 1999.

[4] M. G. Siqueira and A. Alwan, "Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids," *IEEE Transactions on speech and audio processing*, vol. 8(4), pp. 443–453, 2000.

[5] M. A. Stone and B. C. J. Moore, "Tolerable hearing aid delays. i. estimation of limits imposed by the auditory path alone using simulated hearing losses," *Ear and Hearing*, vol. 20(3), pp. 182–191, 1999.

[6] H. A. L. Joson, F. Asano, Y. Suzuki, and S. Toshio, "Adaptive feedback cancellation with frequency compression for hearing aids," *Journal of the Acoustical Society of America*, vol. 94(6), pp. 3248C3254, 1993.

[7] J. L. Nielsen and U. P. Svensson, "Performance of some linear time-varying systems in control of acoustic feedback," *Journal of the Acoustical Society of America*, vol. 1(106), pp. 240–254, 1999.

[8] C. Boukis, D. P. Mandic, and A. G. Constantinides, "Toward bias minimization in acoustic feedback cancellation systems," *Journal of the Acoustical Society of America*, vol. 121(3), pp. 1529–1237, 2007.

[9] H. R. Skovgaard, "Hearing aid compensating for acoustic feedback," in *Patent, US*, 1997, p. 5680467.

[10] J. Hellgren and U. Forssell, "Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 9(8), pp. 906–913, 2001.

[11] N. A. Shusina and B. Rafaely, "Feedback cancellation in hearing aids based on indirect close-loop identification," in *Proc. IEEE Benelux Signal Process. Symp*, 2002, pp. 177–180.

[12] A. Spriet, I. Proudler, M. Moonen, and J. Wouters, "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," *IEEE Transactions on Signal Processing*, vol. 53(10), pp. 3749–3763, 2005.

[13] B. S. Atal, "The history of linear prediction," *IEEE Signal Processing Magazine*, vol. 23(2), pp. 154–161, 2006.

[14] T. E. Tremain, "The government standard linear predictive coding algorithm: Lpc10," *Speech Technology*, vol. 1, pp. 40–49, 1982.

[15] J. R. Deller, J. G. Proakis, and J. H. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Company, 1993.

[16] S. Haykin, *Adaptive Filter Theory, 4th edition*, Prentice-Hall, 2002.

[17] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*, Prentice-Hall, 1978.

[18] N. Levinson, "The wiener rms (root-mean-square) error criterion in filter design and prediction," *Journal of Mathematical Physics*, vol. 25, pp. 261–278, 1947.

[19] J. Durbin, "The fitting of time series models," *Review of the Institute for International Statistics*, vol. 28, pp. 233–243, 1960.

[20] J. P. Burg, "Maximum entropy spectral analysis," in *Proceedings of the 37th Meeting of the Society of Exploration Geophysicists*, 1967.

[21] H. F. Chi, S. X. Gao, S. D. Soli, and A. Alwan, "Band-limited feedback cancellation with a modified filtered-x lms algorithm for hearing aids," *Speech Communication*, vol. 39(1), pp. 147–161, 2003.

[22] J. M. Kates, *Feedback cancellation in a hearing aid with reduced sensitivity to low-frequency tonal inputs*, US Patent, US 6831986, 2004.

[23] B. Widrow and S. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

[24] E. Bjarnason, "Analysis of the filtered-x lms algorithm," *IEEE Transactions on Speech and Audio Processing*, vol. 3(6), pp. 504–514, 1995.

[25] M. M. Yasmin and M. B. Jose Carlos, "Mean weight behavior of coupled lms adaptive systems applied to acoustic feedback cancellation in hearing aids," in *Proc. ICISP*, 2008, pp. 527–535.

[26] Digital Signal Processing Committe, *Programs for Digital Signal Processing*, IEEE Press, New York, 1979.

[27] L. H. Zetterberg and Q. Zhang, "Elimination of transients in lpc based vocoders," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, 1988, pp. 1079–1082.

[28] M. G. Hall, A. V. Oppenheim, and A. S. Willsky, "Time-varying parametric modeling of speech," *Signal Processing*, vol. 5(3), pp. 267–285, 1983.

Paper III

Extracting the Invariant Model from the Feedback Paths of Digital Hearing Aids

G. Ma, F. Gran, F. Jacobsen and F. T. Agerkvist

Submitted to Journal of the Acoustical Society of America, April 2010

Extracting the invariant model from the feedback paths of digital hearing aids

Guilin Ma[a)] and Fredrik Gran
Research Group, GN ReSound A/S, Lautrupbjerg 7, 2750 Ballerup, Denmark

Finn Jacobsen and Finn Agerkvist
Acoustic Technology, Department of Electrical Engineering, Building 352, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

(Dated: April 29, 2010)

Feedback whistling is a severe problem with hearing aids. A typical acoustical feedback path represents a wave propagation path from the receiver to the microphone and includes many complicated effects, among which some are invariant or nearly invariant for all users and in all acoustical environments given a specific type of hearing aids. Based on this observation, a feedback path model that consists of an invariant model and a variant model is proposed. A common-acoustical-pole and zero model-based approach and an iterative least-square search-based approach are used to extract the invariant model from a set of impulse responses of the feedback paths. A hybrid approach combining the two methods is also proposed. The three methods are verified using the artificial datasets and cross-validated using the measured feedback paths. The results show that the proposed hybrid method gives the best overall performance, and the extracted invariant model is effective in modeling the feedback path.

PACS numbers: 43.20.El, 43.60.Ac, 43.60.Uv

I. INTRODUCTION

Feedback is one of the most severe problems with hearing aids. It refers to a process in which a part of the receiver output is picked up by the microphone, amplified by the hearing-aid processing and sent out by the receiver again. When the hearing-aid amplification is larger than the attenuation of the feedback path (FBP), instability occurs and usually results in feedback whistling, which limits the maximum gain that can be achieved[1] and compromises the comfort of wearing hearing aids.

Although electrical and mechanical FBPs also exist, the acoustical FBP is found to be the most significant contributor to the feedback signal.[2] A typical acoustical FBP represents a wave propagation path from the receiver to the microphone and includes the effects of the receiver, the hook, the tube, the fitting of the earmold, the vent, the pinna, the external acoustics, the microphone, etc. For a specific type of hearing aids, these complicated effects generally fall into two categories: the first category is invariant or varying very slowly for all users and in all acoustical situations. This includes the responses of the microphone, the receiver, the hook, etc. The second category is dependent on the individual characteristics, such as the length of PVC tube, the shape of earmold and ear canal, etc., or on the change of the hearing-aid fitting and the external acoustical environment. It is the second category that results in a large inter-subject variability of the FBP and also a large variation of the FBP over time. The properties of these two kinds of effects have been investigated intensively in the previous studies.[3,2,4]

The modeling of the FBP has been addressed from two perspectives. Some studies have focused on the equivalent electro-acoustical models of cavities, transducers, etc.[5,6,7] However, it is very difficult for these models to be precise due to oversimplification of the problem or unknown/imprecise parameters. The other studies have focused on the FBP model used in the adaptive feedback cancellation (AFC), which is an approach widely used for suppressing feedback.

The traditional AFC uses an adaptive Finite-Impulse-Response (FIR) filter to model the overall FBP.[8] This model needs a long filter to cover the major part of the FBP impulse response and therefore has a slow converge speed and a high computational load. To address these issues, an alternative form of the FBP model has been proposed, which represents the FBP with two parts: a short adaptive FIR filter and a fixed filter (usually an IIR filter).[9] The fixed filter aims at modeling the invariant or slowly-varying portion of the FBP, whereas the adaptive filter tracks the rapidly-changing part. This model generally yields a shorter adaptive FIR filter, a faster converge speed and a smaller computational load. However, the way to obtain the coefficients of the fixed filter in practice is to measure the FBP for each individual user when the hearing aid is fitted to the user, and fit the fixed filter so that it models the measured response. This not only requires an additional fitting step, but also fails to capture the true invariant part of the FBP because the measured FBP already includes some of the variant characteristics. This paper presents the methods to extract the invariant part of the FBP. Since the invariant part is independent of the users and the acoustical environ-

[a)] Also at Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark; Electronic address: mguilin@gnresound.dk, gm@elektro.dtu.dk

ments, the new methods can avoid the individual fitting procedure. When the extracted part is represented in the fixed filter of the FBP model, the adaptation of AFC mainly on the variant part is truly possible.

The basic idea of the extraction is to measure a number of FBPs for the same type of hearing aids but on different users, and then extract the part that is common in all the measured FBPs. This resembles the problem of extracting the source and receiver independent part of the room transfer function from a set of room transfer functions measured in the same room, and the problem of extracting the direction independent part of the head-related transfer function from a set of head-related transfer functions measured with the same head. To address these two problems, two methods have been proposed. The first method is based on the common-acoustical-pole zero (CPZ) model, which assumes that the part to be extracted contains only poles.[10,11] The second method, on the contrary, makes no assumption on the pole-zero structure of the part to be extracted and uses an iterative least-square search (ILSS) to find its impulse response.[12] In this paper, the CPZ model-based method and the ILSS approach are both used to extract the invariant part of the FBP. A combination of the two methods is also proposed and found to give the best performance.

The outline of the paper is as follows: in Section II two FBP models used in AFC are explained and discussed. In Section III, the CPZ model and the ILSS approach are described. A hybrid approach that combines the two methods is proposed. Then the results of extracting the invariant part of the FBP using artificial data are presented in Section IV. In Section V the FBP measurement configuration and procedure are described, and the results using the measured data are presented. In the end, concluding remarks and directions for future work are given in Section VI.

II. FEEDBACK PATH MODELS IN THE ADAPTIVE FEEDBACK CANCELLATION

The general diagram of the AFC is depicted in Fig. 1. As mentioned in Section I, many complicated effects are included in the FBP impulse response $b(n)$, where n denotes the discrete-time index and starts from 0. The impulse response of the FBP model $\hat{b}(n)$ is adaptive in order to track the variations in the FBP.

Due to the problem of computational complexity, instability and local minima with the adaptive Infinite-Impulse-Response (IIR) filter,[8] the traditional FBP model $\hat{b}(n)$ used in the AFC is simply an adaptive FIR filter, which does not distinguish the invariant part from the variant part of the FBP but models the overall FBP with a single FIR filter. Therefore it usually needs a long filter and adapts on the whole model $\hat{b}(n)$. This results in a slow converge speed and a heavy computational load.

The alternative FBP model in the AFC decomposes $\hat{b}(n)$ into two parts as shown in Fig. 2. The first part is a fixed filter (usually an IIR filter) with the impulse response $d(n)$, which approximates the FBP measured

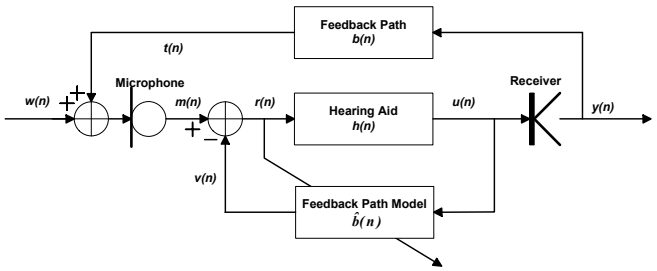


FIG. 1. General diagram of the AFC. The input to the hearing-aid processing is $m(n)$, which is the sum of desired input signal $w(n)$ and the feedback signal $t(n)$. The processed hearing-aid signal is $u(n)$. The signal output into the ear canal is $y(n)$. The impulse response of the FBP is $b(n)$, and $v(n)$ is the instantaneous estimate of $t(n)$, which is subtracted from $m(n)$ to remove the feedback. The impulse response of the modeled FBP is $\hat{b}(n)$. It should be noted that the impulse responses of the microphone and receiver have been included in $b(n)$.

when the hearing aid is fitted to the user. The second part is an on-line adaptive FIR filter with the impulse response $c(n)$ modeling the changes of the FBP over time.[9] The idea of this decomposition is to take out the invariant or slowly-varying part of the FBP as much as possible from the adaptive model and use a short adaptive filter to track the rapidly-varying portion, which primarily relates to the changing external acoustics. Compared with the previous FBP model, this approach yields a shorter adaptive FIR filter, a faster converge speed and a smaller computational load. However, this model requires an additional procedure to measure the FBP using a probe signal for each individual user. Besides, the obtained fixed filter representing the measured FBP includes not only the invariant effects but also some variant effects. For example, the fitting of the hearing aid in the ear canal is captured in $d(n)$ but it is subject to changes when the user yawns or when the hearing aid is re-inserted to the ear.

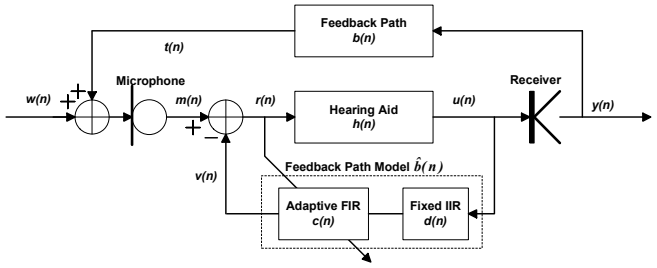


FIG. 2. The diagram of the AFC with an alternative FBP model, which divides the impulse response of the FBP model $\hat{b}(n)$ into two parts: a fixed IIR filter with the impulse response $d(n)$ obtained through the measurement of the true FBP and a short adaptive FIR filter with the impulse response $c(n)$ used to track the variations of the FBP.

To achieve an ideal decomposition of the FBP model and avoid individual measurement, the fixed filter should only include the true invariant part that is independent of the users and the acoustical environments, as illustrated

in Fig. 3. The details on how to extract the invariant model $f(n)$ are given in Section III.

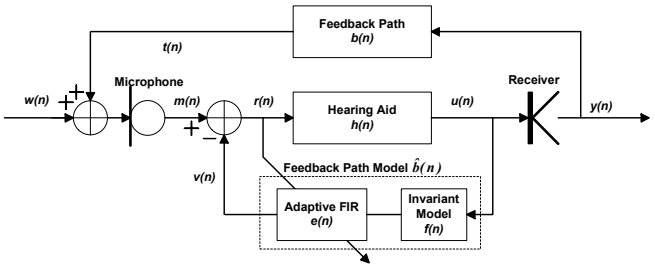


FIG. 3. The diagram of the AFC with a proposed FBP model, which divides the impulse response of the FBP model $\hat{b}(n)$ into two parts: the invariant FBP model, which could be either an FIR or an IIR filter with the impulse response $f(n)$ and the variant FBP model, which is a short adaptive FIR filter with the impulse response $e(n)$ to track the variations of the FBP.

It should be noted that the invariant part of the FBP is only meaningful for the same type of hearing aids. In reality, the invariant part of the FBP is not strictly invariant from device to device due to the variation within the batch of components. It will be investigated in Section V whether this variation is small enough so that the modeling of the invariant part still makes sense in practice.

III. METHODS FOR EXTRACTING THE INVARIANT FEEDBACK PATH MODEL

The extraction of the invariant part of the FBP can be done in several ways. The first way is to measure it directly. However, since in practice the invariant part is coupled with the variant part very closely, it will be very difficult to isolate the invariant part unless each component is detached from the hearing aid and measured individually, which requires high precision in the measurements. Furthermore, the measured invariant part is only valid for a single device due to the variation within the batch of components. The second way is to model each component either theoretically by using an equivalent electro-acoustical model[5,6,7] or numerically by using methods such as boundary element calculations.[4] To yield a good estimate of the invariant part, these methods need to build a precise model for every component, which may be difficult for some of the components. The third way is to extract the invariant part from a set of measured FBPs as has been done for room transfer function and head-related transfer function modeling. This paper focuses on this approach. The idea is to measure a number of FBPs using the same type of hearing aids on different users. The invariant part of the FBP can then be regarded as the common part of these measured FBPs.

Suppose L FBPs have been measured with the impulse responses $b_1(n), b_2(n), \cdots, b_L(n)$. In principle, the FBP impulse responses have infinite duration. It is assumed in the following that the impulse responses of the FBP and the FBP models are all truncated to a sufficient length N so that the energy loss in the impulse responses due to the truncation is at least 35 dB below the total energy of the true responses.

Let $f(n)$ and $e_k(n)$ denote the impulse response of the invariant model and the variant model of the k-th FBP respectively. The k-th modeled FBP $\hat{b}_k(n)$ is then the convolution of $e_k(n)$ and $f(n)$, i.e.

$$\hat{b}_k(n) = e_k(n) \odot f(n), \tag{1}$$

where $\odot$ is the convolution operator, and the symbol $\hat{}$ is used to denote the estimate of the corresponding quantity in this paper.

One way to formulate the extraction problem is to estimate $f(n)$ with the objective of minimizing the difference between the the modeled FBP $\hat{b}_k(n)$ and the true FBP $b_k(n)$. Due to the different vent sizes, pinna shapes and microphone locations for different users, some of the measured FBP impulse responses may contain more energy than others. This will result in a preference of minimizing the modeling error for large FBPs. If the measurement is conducted in the same way for all the measured FBPs, every measured FBP should be treated equally. Therefore, the measured impulse responses $b_k(n)$ is first scaled in the following way:

$$\tilde{b}_k(n) = b_k(n) \times g_k, \tag{2}$$

$$g_k = \frac{1}{\sum_{i=0}^{N-1} |b_k(i)|^2}, \tag{3}$$

so that $\sum_{i=0}^{N-1} |\tilde{b}_k(i)|^2$ is unity for any k. The extraction problem can then be formulated as follows:

$$\hat{f}(n) = \arg\min_{f(n)} \|\tilde{\mathbf{B}} - \hat{\mathbf{B}}\|_2^2, \tag{4}$$

$$\tilde{\mathbf{B}} = \left[\ \tilde{\mathbf{b}}_1^T,\ \ldots\ , \tilde{\mathbf{b}}_L^T\ \right]^T, \tag{5}$$

$$\hat{\mathbf{B}} = \left[\ \hat{\mathbf{b}}_1^T,\ \ldots\ , \hat{\mathbf{b}}_L^T\ \right]^T, \tag{6}$$

$$\tilde{\mathbf{b}}_k = \left[\ \tilde{b}_k(0),\ \ldots\ , \tilde{b}_k(N-1)\ \right]^T, \tag{7}$$

$$\hat{\mathbf{b}}_k = \left[\ \hat{b}_k(0),\ \ldots\ , \hat{b}_k(N-1)\ \right]^T, \tag{8}$$

where $\|\|_2$ denotes the Euclidean norm, the superscript T denotes the transpose of a matrix or a vector, and $\hat{b}_k(n)$ is defined in Eq.(1). The bold symbol represents a matrix or a vector.

This optimization problem is non-linear. The solution methods based on the CPZ model and the ILSS method are described below.

There are also other ways to formulate the extraction problem besides the one in Eqs.(4-8). For example, the problem can be formulated in the frequency domain and a weighting for the importance of each frequency bin can be applied on the optimization problem. This will result in corresponding changes in the following solution methods.

A. Common-acoustical-pole and zero modeling

The CPZ model was first proposed to model the room transfer function, and later it was used to model the

head-related transfer function.[10,11] It uses the poles to model the resonances in the room or in the ear canal, which are common to all the room transfer functions in the same room or all the head-related transfer functions for the same head.

For FBP modeling, the invariant part includes the responses of the receiver, the tube inside the hearing aid shell, the hook, the microphone, etc., most of which also exhibit resonances. Therefore, it should also contain common poles although common zeros may also exist. Since resonances usually need long FIR filters to model, the CPZ model should capture the majority of the invariant part of the FBP if the number of common zeros is not very large. In this case, the small number of common zeros can be moved to the short FIR filter in the variant model $e_k(n)$.

To estimate the common poles, a number of measured impulse responses should be used instead of one single impulse response because poles are strongly affected or canceled by zeros in a single impulse response.[10]

In the CPZ model-based approach for extracting the invariant part of the FBP, each FBP is modeled by an Autoregressive Moving Average (ARMA) model:

$$\hat{b}_k(n) = -\sum_{i=1}^{P} a_i \hat{b}_k(n-i) + \sum_{i=0}^{Q} c_{i,k}\delta(n-i), \qquad (9)$$

where δ is the unit pulse function ($\delta(n) = 1$ for $n = 0$, and $\delta(n) = 0$ for any other n), a_i's are the coefficients of the Autoregressive (AR) model, which is common to all the FBP models, and $c_{i,k}$'s are the coefficients of the Moving Average (MA) model for the k-th FBP model. In this ARMA model, the AR model with P poles models the invariant part of the FBP, and the MA model with Q zeros (which may include common zeros) models the variant part. The impulse responses $f(n)$ and $e_k(n)$ then correspond to the impulse response of the common AR model and the MA model of the k-th FBP model respectively. The estimation of $f(n)$ in Eq.(4) becomes an estimation of a_i's

$$\{\hat{a}_i\}_{i=1}^{P} = \arg\min_{a_1,\ldots,a_P} \|\tilde{\mathbf{B}} - \hat{\mathbf{B}}\|_2^2, \qquad (10)$$

which is known to be difficult.[13] However it can be reformulated as a new problem, by replacing the error between the modeled FBP and the actual FBP with the so-called "equation error". The optimal analytic solution to the new problem exists although it can be suboptimal to the original problem in Eq.(10),[10]

$$x = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\overline{\mathbf{B}}, \qquad (11)$$

$$x = \left[\hat{\mathbf{a}}^T, \hat{\mathbf{c}}_1^T, \ldots, \hat{\mathbf{c}}_L^T\right]^T, \qquad (12)$$

$$\hat{\mathbf{a}} = \left[-\hat{a}_1, \ldots, -\hat{a}_P\right]^T, \qquad (13)$$

$$\hat{\mathbf{c}}_k = \left[\hat{c}_{0,k}, \ldots, \hat{c}_{Q,k}\right]^T, \qquad (14)$$

$$\overline{\mathbf{B}} = \left[\ \overline{\mathbf{b}}_1^T, \ldots, \overline{\mathbf{b}}_L^T\ \right]^T, \qquad (15)$$

$$\overline{\mathbf{b}}_k = \left[\ \tilde{b}_k(0), \ldots, \tilde{b}_k(N-1)\ , \mathbf{0}_{1\times P}\right]^T, \qquad (16)$$

where $\hat{a}_i$'s and $\hat{c}_{k,i}$'s are the estimate of a_i's and $c_{k,i}$'s respectively, $\mathbf{0}_{1\times P}$ is a row vector containing P zeros and the matrix $\mathbf{A}$ is defined in Appendix A.

B. Iterative least-square search

The ILSS method was proposed for approximating the greatest common divisor of a polynomial[14] and later used to model the head-related transfer function.[12] It finds the common impulse response among a set of impulse responses using an iterative least-square search.

As mentioned in Section III.A, the invariant model of FBP may contain not only poles but also zeros. Therefore, the ILSS approach, which does not make assumptions on the pole-zero structure but estimates the impulse response directly, should be more general.

Suppose that the length of the impulse response of the invariant model $f(n)$ and the variant model $e_k(n)$ is truncated to C and M respectively, and that $M+C-1 \le N$. The FBP model $\hat{b}_k(n)$ of the length N is then the convolution between $e_k(n)$ and $f(n)$ with zero-padding:

$$\hat{\mathbf{b}}_k = \left[\mathbf{e}_k^T\mathbf{F}, \mathbf{0}_{1\times(N+1-M-C)}\right]^T, \qquad (17)$$

$$= \left[\mathbf{f}^T\mathbf{E}_k, \mathbf{0}_{1\times(N+1-M-C)}\right]^T, \qquad (18)$$

$$\mathbf{f} = \left[f(C-1), f(C-2), \ldots, f(0)\right]^T, \qquad (19)$$

$$\mathbf{e}_k = \left[e_k(M-1), e_k(M-2), \ldots, e_k(0)\right]^T, \qquad (20)$$

where $\mathbf{0}_{1\times(N+1-M-C)}$ is a row vector with $(N+1-M-C)$ zeros, the convolution matrix $\mathbf{E}_k$ and $\mathbf{F}$ are formed by $e_k(n)$ and $f(n)$ respectively and defined in Appendix B.

To obtain the estimate of $f(n)$, an iterative search is performed in four steps:[12]

Step 1: Set iteration counter $s = 0$, and set $\hat{\mathbf{f}}$ to an initial value $\hat{\mathbf{f}}^0$, where the superscript denotes the iteration number and the symbol $\hat{}$ denotes the estimate of the corresponding quantity at that iteration.

Step 2: Given $\hat{\mathbf{f}}^s$, the least-square solution to the optimization problem

$$\{\hat{\mathbf{e}}_k^s\}_{k=1}^{L} = \arg\min_{\mathbf{e}_1,\ldots,\mathbf{e}_L} \|\tilde{\mathbf{B}} - \hat{\mathbf{B}}\|_2^2, \qquad (21)$$

is

$$\left[\ \hat{\mathbf{e}}_1^s \ldots, \hat{\mathbf{e}}_L^s\ \right] = \left(\hat{\mathbf{F}}^s\left(\hat{\mathbf{F}}^s\right)^T\right)^{-1}\hat{\mathbf{F}}^s\tilde{\mathbf{B}}_1, \qquad (22)$$

where

$$\tilde{\mathbf{B}}_1 = \left[\tilde{\mathbf{b}}_1^{tr}, \ldots, \tilde{\mathbf{b}}_L^{tr}\right], \qquad (23)$$

$$\tilde{\mathbf{b}}_k^{tr} = \left[\tilde{b}_k(0), \ldots, \tilde{b}_k(M+C-2)\right]^T, \qquad (24)$$

where the superscript tr stands for truncation of the matrix or vector.

Step 3: Given $\hat{\mathbf{e}}_k^s$, the least-square solution to the optimization problem

$$\hat{\mathbf{f}}^{s+1} = \arg\min_{\mathbf{f}} \|\tilde{\mathbf{B}} - \hat{\mathbf{B}}\|_2^2, \qquad (25)$$

is

$$\hat{\mathbf{f}}^{s+1} = \left(\hat{\mathbf{E}}^s \left(\hat{\mathbf{E}}^s\right)^T\right)^{-1} \hat{\mathbf{E}}^s \tilde{\mathbf{B}}_2, \tag{26}$$

where the matrix $\mathbf{E}$ is defined in Appendix B, and

$$\tilde{\mathbf{B}}_2 = \begin{bmatrix} \tilde{\mathbf{b}}_1^{tr} \\ \vdots \\ \tilde{\mathbf{b}}_L^{tr} \end{bmatrix}. \tag{27}$$

Step 4: $s = s + 1$, and repeat *Step 2* and *Step 3* until s reaches a predetermined value.

The ILSS method has proved to converge to a local minimum.[14] The initial value might be of importance in the search of good estimates.

C. Hybrid approach

The CPZ model-based approach and the ILSS method can be combined to yield a new hybrid approach, which is referred to as the "ILSSCPZ" method in this paper. The ILSSCPZ method consists of two steps: (1) the invariant model is extracted using the CPZ model-based approach described in Section III.A; (2) the impulse response of the extracted AR model is truncated and used as an initial estimate in the ILSS method, and the four steps described in Section III.B are performed. Since the ILSS method can be trapped in a local minimum, the ILSSCPZ may provide a better converged solution by using a good initial estimate obtained from the CPZ model-based approach.

IV. RESULTS USING ARTIFICIAL FEEDBACK PATHS

The invariant part is very difficult to measure in practice. Therefore to verify that the methods described in Section III can extract the invariant part of the FBPs effectively, and to investigate the properties of each method, artificial FBPs are first used.

A. Artificial data generation

An impulse response of the FBP was measured with a hearing aid in free space. The measurement procedure is described in Section V.A. The leading delay of the impulse response was first removed and then the impulse response was fitted by an ARMA model with 4 poles and 7 zeros. The impulse responses of the measured FBP and the fitted ARMA model are illustrated in Fig. 4.

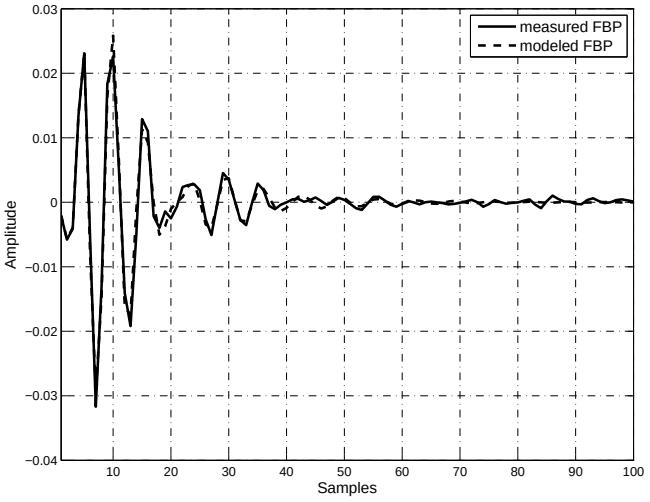


FIG. 4. The impulse responses of the measured FBP and the fitted ARMA model.

The fitted ARMA model is then regarded as the invariant part of the FBPs. Its impulse response is truncated to 255 samples and used to generate three types of artificial FBPs. For each type of FBPs, 1000 FBPs are generated.

Type I: Assume that the variant part of the FBP is a 50-tap FIR filter with coefficients following the standard normal distribution (zero mean and unit variance). Each artificial FBP is computed as the convolution of the FIR filter and the truncated impulse response of the ARMA model amplified by a uniformly distributed random gain between 0 and 10 dB, which can cover the range of the level difference in the measured FBPs in most cases. This dataset is mainly used to test the effectiveness of the methods with random variant parts and different gain levels.

Type II: The artificial FBPs are generated by delaying the truncated impulse response of the ARMA model randomly between 0 and 3 samples and amplifying the delayed impulse response with a uniformly distributed random gain between 0 and 10 dB. When the sampling frequency is 15625 Hz, a delay of 3 samples corresponds to a spatial delay of $1/15625 * 3 * 343$ m ≈ 0.066 m, where 343 m/s is the speed of sound at 20°C. This spatial delay is large enough to cover the range of the difference in the feedback propagation lengths caused by different sizes of the ear. The delays for this type of artificial FBPs are all rounded to the nearest integers, and therefore this dataset is primarily used to test the three methods in terms of the sensitivity of integer delays.

Type III: The artificial FBPs are generated in the same way as for dataset *Type II*, but the delays are not restricted to integers. This dataset is mainly used to test the three methods in terms of the sensitivity of fractional delays.

In order to test the noise sensitivity of the three methods, three more datasets were generated by adding -30 dB noise to the three clean datasets described above.

B. Simulation results

For each type of artificial FBPs, the 1000 FBPs are divided into 100 groups, each including 10 FBPs. The three methods extract the invariant part from every group of FBPs. Therefore, for every type of FBPs, 100 estimates were obtained using each method, and used to show the variance of the estimation.

For all the methods, each group of the artificial FBPs is first normalized (cf. Eqs.(2)-(3)) so that the 10 FBPs are of the same energy. Then the procedures described in Section III are followed to extract the invariant part.

For dataset *Type I*, the CPZ model-based approach uses the true number of poles and zeros, i.e., 4 poles and 56 zeros (49 from the variant part, and 7 from the invariant ARMA model). For the ILSS method, M, the length of the impulse response of the variant model, is set to the correct length of 50 samples. However, it was found that C, the length of the impulse response of the invariant model, should not be as large as 255 samples because the "active" region of the impulse response of the ARMA model is not very wide (cf. Fig.4). An excessively large C will give too many degrees of freedom to the optimization and lead to strange estimates. In the simulation C is set to 60 samples. The ILSSCPZ method simply uses the parameters in the CPZ model-based approach and the ILSS method.

For dataset *Type II*, the CPZ model-based approach still uses 4 poles but 10 zeros (7 from the invariant ARMA model) to allow 3 zeros to compensate for the integer delay between 0 and 3 samples. For the ILSS method, M is also 4 samples corresponding to 3 zeros and C is set to 60 samples as for the dataset *Type I*. The ILSSCPZ method uses the parameters in the CPZ model-based approach and the ILSS method.

For dataset *Type III*, the CPZ model-based approach still uses 4 poles but the number of zeros is not easy to determine since the fractional delay usually needs a certain number of zeros to represent. It was found that 16 zeros (7 from the invariant ARMA model) to allow 9 zeros to compensate for the fractional delay between 0 and 3 samples gave the best performance. For the ILSS method, it was found that $M = 4$ gave the best results when C is set to 60 samples. The reason why M has to be small when C is 60 samples for the ILSS method is because the excess in the total degrees of freedom will degrade the estimates very much. The ILSSCPZ method uses the parameters in the CPZ model-based approach and the ILSS method.

It should be noted that the ILSS method and the ILSSCPZ method have inherent ambiguities in the overall power level and leading delay of the estimates because the variant model can absorb these ambiguities. This is not a problem for the AFC purpose, but to show the performance of the methods, the estimates from these two method are aligned and normalized.

In the simulation, different initial estimates $\hat{\mathbf{f}}^0$ were tried for the ILSS method, including the all-one sequence, the random sequence, and the truncated average impulse response of the 10 FBPs in the group. The difference resulted from these different initial estimates is very small, which agrees with the finding in the study of the head-related transfer function modeling[12]. Therefore only the results using the all-one sequence as the initial estimate are presented.

The results for the three methods using the three clean datasets and three noisy datasets are illustrated in Figs. 5-8. As can be seen from Fig. 5, the CPZ model-based approach can estimate the common poles very successfully when there is no noise in the datasets. When noise is present, the estimate error is large. The impulse responses of the mean poles in Fig. 6 indicate that underestimation of the pole response occurs when the noise is present. This shows that the CPZ model-based approach is sensitive to noise. The reason is that the analytic solution in Eqs.(11-16) is not necessarily optimal for the optimization problem in Eq.(10), which is known to be difficult as mentioned in Section III.A.

The ILSS method with all-one initialization generally provides good estimates of the invariant part and is not sensitive to noise. However, for dataset *Type I*, there are larger variances due to the local minima.

The ILSSCPZ method yields good estimates for all the datasets. Compared with the ILSS method with all-one initialization, the variances for dataset *Type I* are reduced.

V. RESULTS USING MEASURED FEEDBACK PATHS

In this section, the extraction performance of the methods is evaluated using the measured FBPs.

A. Measurement of the feedback paths

Ten FBPs were measured with ten commercial BTE devices of the same type, Beltone ACCESS 75D with ThinTube. The ten hearing aids were put on ten users including five men and five women.

The impulse responses of the FBPs were computed by cross-correlating the probe signal sent out through the receiver and the recorded microphone signal. The probe signal is the maximum-length sequence (MLS) with a period of 255 samples. The MLS is repeated for a number of times to obtain a high Signal-to-Noise Ratio (SNR), around 30 dB, for the FBP response relative to random ambient noise. The sampling frequency was 15625 Hz. The detailed measurement procedure can be found in Kates 2001.[15] All the impulse responses are of a length of 255 samples, which are shown in Fig. 9.

It should be noted that the variation within the batch of components of digital hearing aids has been included in the measured FBPs since ten different devices of the same type were used.

B. Simulation results

Since the true invariant part of the measured FBPs is not available in practice, a direct evaluation of the methods is difficult. However, the ultimate objective of extracting the invariant part is to achieve a good FBP model. Therefore, the effectiveness of the extraction can be assessed by fitting the measured FBP using the estimated invariant FBP model.

To avoid overfitting in the invariant FBP modeling and to make the best use of the limited number of the measured FBPs, the leave-one-out cross validation is adopted in this paper,[16] which extracts the invariant part based

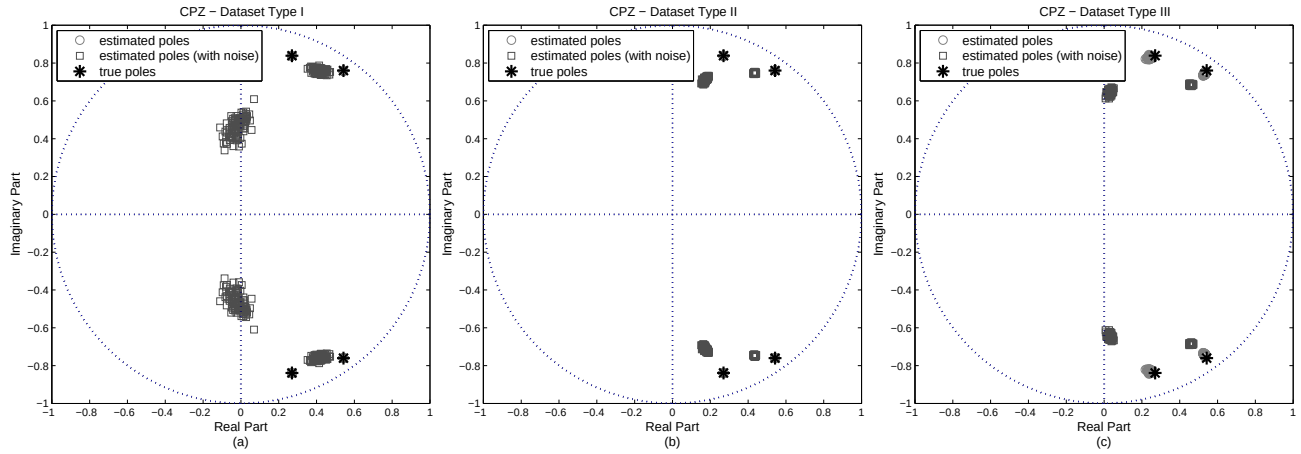


FIG. 5. The estimated poles of the invariant model using the CPZ model-based approach with and without noise in the datasets.

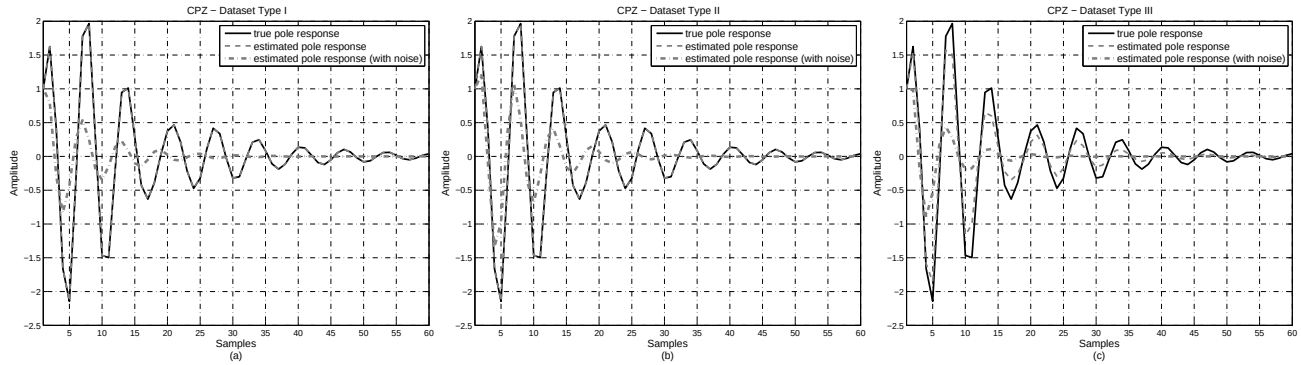


FIG. 6. The impulse responses of the mean poles of the invariant model estimated using the CPZ model-based approach with and without noise in the datasets.

on nine out of the ten FBPs, and then uses the extracted invariant model to fit the the remaining one FBP. The fitting is done by estimating the variant model of a given order so that the mean-square error (MSE) of the FBP model is minimized. Assume that the remaining FBP is the k-th FBP. The normalized MSE between the modeled FBP and the measured FBP is calculated as below:

$$J_k = \frac{\|\tilde{\mathbf{b}}_k - \hat{\mathbf{b}}_k\|_2^2}{\|\tilde{\mathbf{b}}_k\|_2^2}. \tag{28}$$

Given a fixed order of the FIR filter in the variant model of the k-th FBP, the quantity J_k can show how well the invariant model is extracted. It has been observed in this study that if the extracted invariant model includes some variant part by mistake, some degrees of freedom of the variant model will be wasted to correct this mistake, and the MSE will generally be larger. Similarly, if the extracted invariant model is not complete, some degrees of freedom have to be used to compensate for the missing invariant part, and the MSE also tends to be larger. A mean normalized MSE is used as a metric to assess the generalized performance of the three methods, which is defined as:

$$J = \frac{1}{L}\sum_{k=1}^{L} J_k, \tag{29}$$

where L equals 10. This metric essentially assembles the information about the invariant model, including the assumptions and the effectiveness. For example, if the variation within the batch of components is so large that the common part of the FBPs is trivial, the order of the variant model has to be very high to reduce J to a certain level. If the extraction of the invariant model by the three methods is not precise enough, the order of the variant model also has to be high to compensate for the estimation error.

To predict the potential benefit of using the proposed model in the AFC, the added stable gain (ASG), which is a widely used metric to assess the performance of the AFC ,[17] is used as the second metric. The ASG for the k-th feedback path I_k and the mean ASG I are also calculated using the leave-one-out cross validation in the

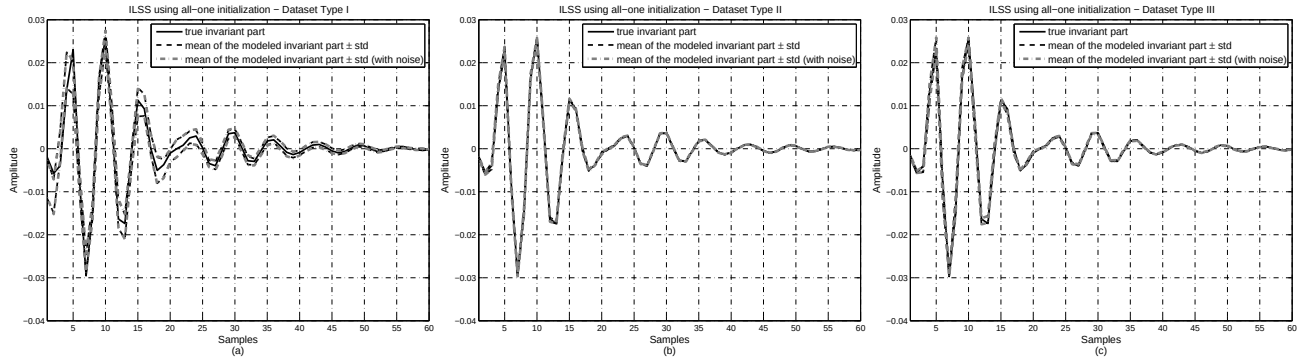


FIG. 7. The impulse responses of the invariant model estimated by the ILSS method using all-one initialization with and without noise in the datasets. The "std" stands for the standard deviation.

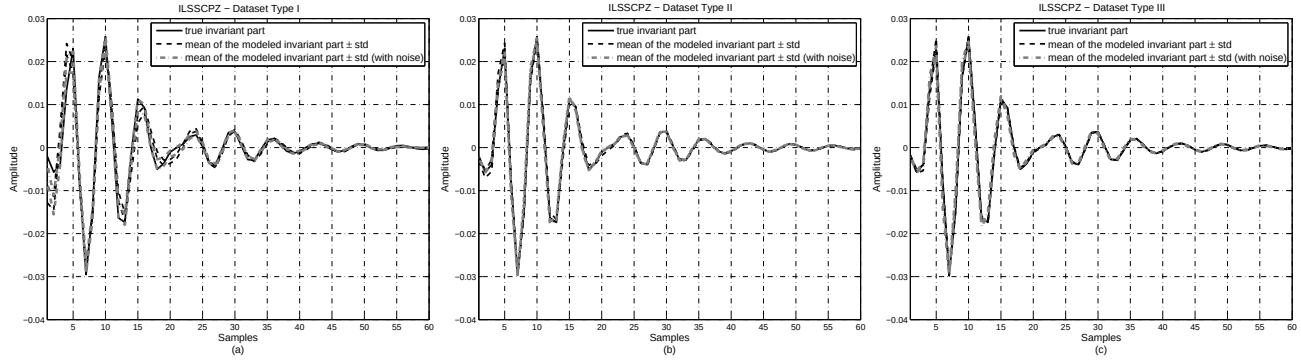


FIG. 8. The impulse responses of the invariant model estimated by the ILSSCPZ method with and without noise in the datasets. The "std" stands for the standard deviation.

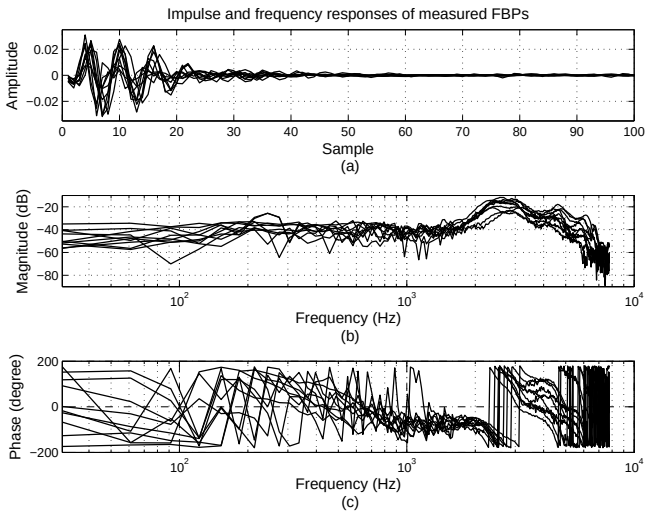


FIG. 9. The impulse responses and frequency responses of the ten measured FBPs.

same way as for the mean normalized MSE:

$$I_k = 20\log_{10}\left(\min_z \frac{1}{|\mathbf{f}_z^H(\tilde{\mathbf{b}}_k - \hat{\mathbf{b}}_k)|}\right) - 20\log_{10}\left(\min_z \frac{1}{|\mathbf{f}_z^H\tilde{\mathbf{b}}_k|}\right), \tag{30}$$

$$\mathbf{f}_z = [\ 1,\ e^{j\omega_z},\ \ldots\ , e^{j\omega_z(N-1)}\ ]^T, \tag{31}$$

$$I = \frac{1}{L}\sum_{k=1}^{L} I_k. \tag{32}$$

where $\omega_z = 2\pi z/N, z = 0, 1, \ldots, N-1$ and the superscript H denotes the conjugate transpose of a matrix or vector. The first term in Eq.(30) indicates the maximum gain the hearing aid can provide when the feedback model is used in the AFC and is determined by the the frequency at which the mismatch between the feedback model and the actual feedback path is the largest. The second term in Eq.(30) indicates the maximum gain the hearing aid can provide without using AFC. Thus the ASG essentially defines the additional gain that can possibly be achieved by using the proposed feedback model in the AFC. It should be noted that in principle the ASG should be calculated using the original impulse responses of the feedback path $b_k(n)$ instead of the scaled responses $\tilde{b}_k(n)$. However, it can easily be shown that this does not influence the ASG calculation because the normalization factor g_k is present in both terms in Eq.(30) and they cancel each other.

The performance of the CPZ model-based approach is illustrated in Fig. 10(a) and Fig. 11(a). As shown in the figures, the larger M the better modeling of the FBPs. In Fig. 10(a), all the curves show a similar trend:

Before the number of poles P reaches 8, the curves are relatively flat or fluctuate. From 8 to 9, all the curves begin to drop, indicating that the performance is improved. Then all the curves become relatively flat again when P increases to a certain point, which is between 9 and 11 depending on M. The curves in Fig. 11(a) are similar to the opposite/negated curves in Fig. 10(a).

For the ILSS method, the initial estimate is an all-one sequence, the same as the one used for the artificial FBPs. The results are illustrated in Fig. 10(b) and Fig. 11(b). Generally speaking, the larger M results in smaller mean normalized MSE and larger mean ASG. The curves are relatively flat when C is larger than 40, but fluctuate when C is smaller than 25 because the ILSS method may converge to a local minimum.

For the ILSSCPZ method, the invariant model is first extracted by the CPZ model-based approach using 11 poles, and then the impulse response of the extracted AR model is truncated to serve as an initial estimate in the ILSS method. The results are shown in Fig. 10(c) and Fig. 11(c). Compared with Fig. 10(b) and Fig. 11(b), for M larger than 13, the performance is improved and the fluctuations as C increases in the curves are much smaller. When C reaches 20, the curves begin to flatten to some extent.

A comparison between of the performance of the CPZ model-based approach and the ILSSCPZ method can be seen in Fig. 10 and Fig. 11. To show the comparison more clearly, the mean normalized MSE and the mean ASG are illustrated as a function of M in Fig. 12. When M is equal to or larger than 16, the performance of the CPZ model-based approach with $P = 9$ and $P = 10$ is close to the performance of the ILSSCPZ method with $C = 10$ and $C = 15$ respectively. When C is larger than or equal to 20, the ILSSCPZ method is better than the CPZ model-based approach with $P = 12$ for almost all M's. Since the performance of the CPZ-model based approach improves very little when P is larger than 11 (cf. Fig. 10 and Fig. 11), the ILSSCPZ method with C larger than or equal to 20 yields better overall performance. This is mainly because the CPZ model-based approach does not model the invariant zeros and is sensitive to noise.

In practice, the variant model usually uses around 10-20 taps in the FIR filter.[18] When M is in the region between 10 and 19, for the CPZ model-based approach with 12 poles, the mean normalized MSE is between -12.0 dB and -16.1 dB and the mean ASG is between 12.8 dB and 17.0 dB. For the ILSSCPZ method with 20 samples of impulse response, the mean normalized MSE is between -12.4 dB and -16.4 dB and the mean ASG is between 13.7 dB and 17.9 dB. These settings are affordable for hearing-aid processors. To achieve greater ASG, the order of the FIR filter M should be increased further.

C. Discussion

The comparison study by J. E. Greenberg *et al.* shows that the the widely used continuously adapting AFC can bring about 8.5 dB ASG on average.[19] A more recent study by A. Spriet *et al.* compares several contemporary AFC approaches and shows that when a 64-tap adaptive FIR filter is used to model the overall FBP, 15 dB ASG can be obtained for speech input signals for all the methods, and that the optimum ASG, which can only be achieved in the ideal situation, is around 16 dB.[20] Compared with the performance reported in these studies, the performance of the proposed model can reach a comparable level with a much shorter adaptive filter. This suggests that the use of the extracted invariant feedback model in the AFC is very promising.

This also indicates that in reality, even when the variation within the batch of components is considered, the invariant part is significant, and the methods used in this paper can effectively extract the invariant FBP model, because if the invariant part is insignificant or if the methods are ineffective, to achieve the same 16 dB ASG as in the previous study,[20] the order of the variant model M would be comparable to 64.

The factors that limit the modeling accuracy of the FBP given a fixed order of the variant model are twofold: Firstly, the methods themselves may converge to local minima. To improve these methods, some heuristic methods can be used to prevent the search from being trapped at local minima easily. Secondly, in practice, both the variation within the batch of components and the individual characteristics are part of the variant model, which need a long FIR filter to model.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, the problem of extracting the invariant model from the feedback paths of digital hearing aids is investigated.

The acoustical feedback path starts from the receiver and ends at the microphone. Along this transmission path, many complicated effects are involved. For a specific type of hearing aids, some of the effects are common/invariant or nearly invariant for all users and in all acoustical environments. Based on this fact, a feedback path model consisting of an invariant model and a variant model is proposed, which differs from the previously proposed model in that it does not require individualized measurements and only identifies the invariant part of the feedback path.

Two extraction methods, i.e. a CPZ model-based approach and an ILSS method, are used to extract the invariant feedback path model based on a set of measured feedback paths. Since the ILSS method suffers from the local minima issue, an ILSSCPZ method is proposed, which uses the CPZ model to provide an initial estimate for the ILSS method. The true invariant part is difficult to obtain in practice, so the three methods are verified using three artificial datasets. The results show that when noise is not present in the datasets, the CPZ model-based approach can estimate the common poles very successfully and when noise is present, it has a large estimation error. The ILSS method is not sensitive to noise and provides good estimates for two out of the three datasets. However, for one of the datasets, the

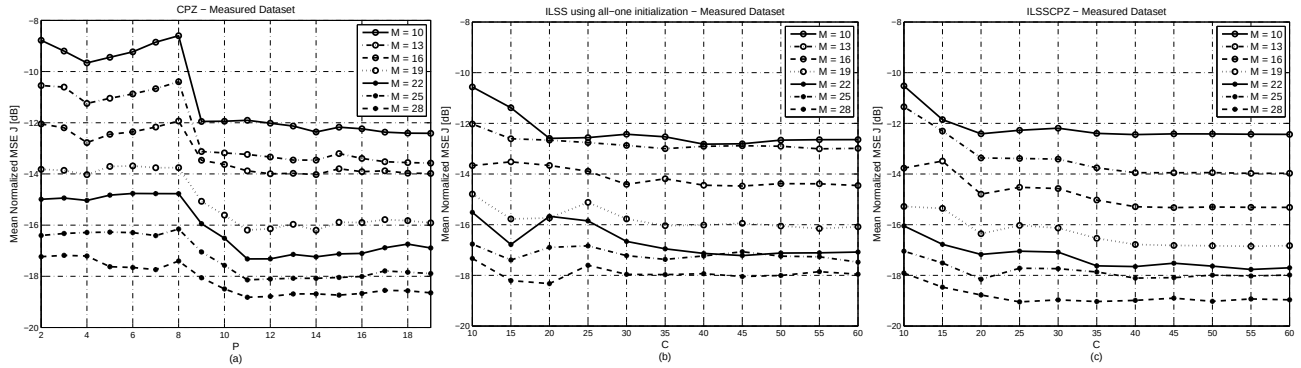


FIG. 10. The cross-validation performance of the methods in terms of mean normalized MSE, where M is the length of the impulse response of the variant model. In Fig. (a), the x-axis is the number of poles P used in the CPZ model-based approach. In Fig. (b) and Fig. (c), the x-axis is the length of the impulse response of the invariant model C used in the ILSS based methods.

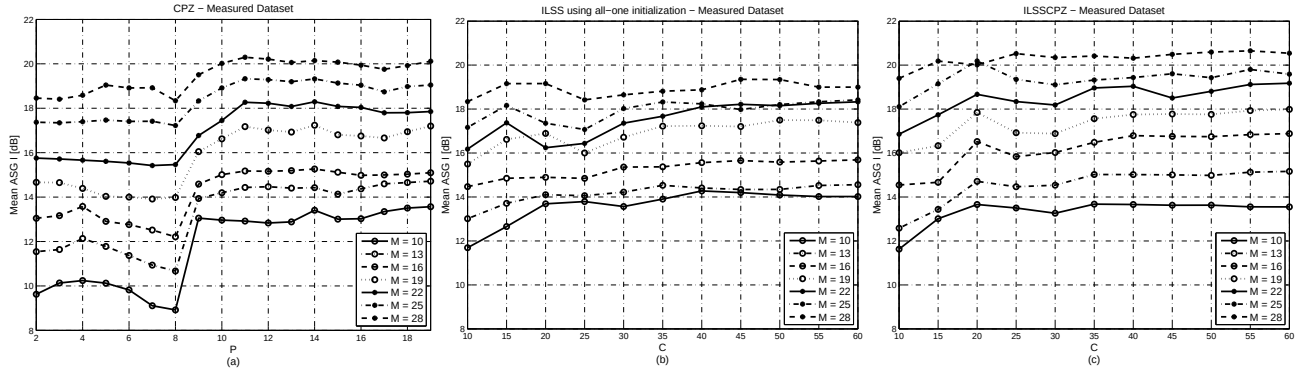


FIG. 11. The cross-validation performance of the methods in terms of mean ASG in dB, where M is the length of the impulse response of the variant model. In Fig. (a), the x-axis is the number of poles P used in the CPZ model-based approach. In Fig. (b) and Fig. (c), the x-axis is the length of the impulse response of the invariant model C used in the ILSS based methods.

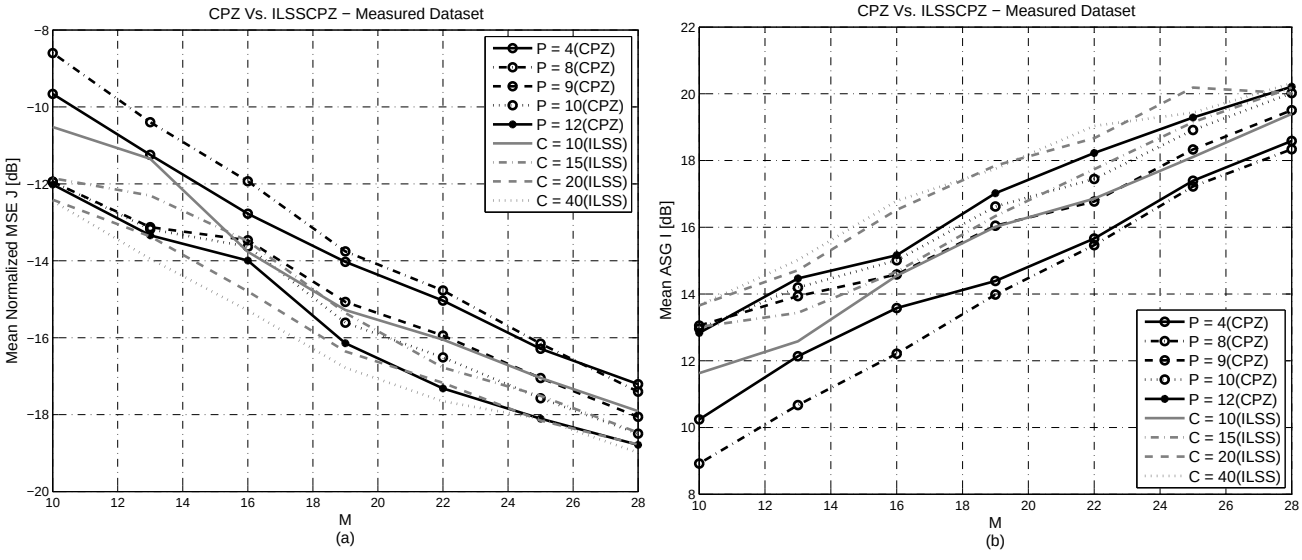


FIG. 12. The performance comparison between the CPZ model-based approach and the ILSSCPZ method in terms of (a) Mean Normalized MSE and (b) Mean ASG.

variances are large. The ILSSCPZ method yields good estimates for all the datasets.

The three methods are then tested on the measured feedback paths. A cross-validation approach is used to investigate whether the invariant part of the feedback paths in practice is trivial and whether the the invariant feedback path model is effective for feedback path modeling. The results show that the proposed ILSSCPZ method gives the best overall performance and the feedback path modeling accuracy improves as the order of the variant model increases. According to the performance achieved by the proposed feedback path model when the order of the variant model is within the range typically used in the hearing aids, the use of the estimated invariant model in the adaptive feedback cancellation is very promising. This also suggests that the invariant part of the feedback path is significant in practice, and that the methods can extract the invariant part effectively. The factors that limit the feedback path modeling accuracy given a fixed order of the variant model are also discussed.

The future work of this study is to investigate the heuristics for the extraction methods that can address the local minima issue and obtain better solutions. Moreover, more types of hearing aids, for example the In-The-Ear (ITE) devices, can be used to verify the methods and the effectiveness of the invariant feedback path model. It is expected that for different types of hearing aids, the effectiveness is different and depends on the proportion of the invariant part relative to the variant part in the overall feedback path.

APPENDIX A: DEFINITION OF THE MATRIX A

The matrix $\mathbf{A}$ used in Eq.(11) is defined as below:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{D} & & \\ \mathbf{A}_2 & & \mathbf{D} & \mathbf{0} \\ \vdots & & & \\ & & \mathbf{0} & \ddots \\ \mathbf{A}_L & & & \mathbf{D} \end{bmatrix}, \tag{A1}$$

where $\mathbf{A}_k$ is of the size $(N+P)\times P$ and defined as:

$$\mathbf{A}_k = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \tilde{b}_k(0) & 0 & \cdots & 0 \\ \tilde{b}_k(1) & \tilde{b}_k(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{b}_k(P-1) & \tilde{b}_k(P-2) & \cdots & \tilde{b}_k(0) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{b}_k(N-1) & \tilde{b}_k(N-2) & \cdots & \tilde{b}_k(N-P) \\ 0 & \tilde{b}_k(N-1) & \cdots & \tilde{b}_k(N-P+1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{b}_k(N-1) \end{bmatrix}, \tag{A2}$$

and $\mathbf{D}$ is of the size $(N+P)\times(Q+1)$ and defined as:

$$\mathbf{D} = \begin{bmatrix} 1 & & & \\ & 1 & \mathbf{0} & \\ & \mathbf{0} & \ddots & \\ & & & 1 \\ 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}. \tag{A3}$$

APPENDIX B: DEFINITION OF THE CONVOLUTION MATRICES

The convolution matrix $\mathbf{F}$ is of the size $M\times(M+C-1)$ and defined as:

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & \cdots & f(C-1) \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & f(0) & \cdots & 0 \\ f(0) & f(1) & \cdots & 0 \end{bmatrix}. \tag{B1}$$

The convolution matrix $\mathbf{E}$ is defined as:

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \vdots \\ \mathbf{E}_L \end{bmatrix}, \tag{B2}$$

where the matrix $\mathbf{E}_k$ is of the size $C\times(M+C-1)$ and defined as:

$$\mathbf{E}_k = \begin{bmatrix} 0 & 0 & \cdots & e_k(M-1) \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & e_k(0) & \cdots & 0 \\ e_k(0) & e_k(1) & \cdots & 0 \end{bmatrix}. \tag{B3}$$

[1] S. F. Lybarger, "Acoustic feedback control," The Vanderbilt Hearing-Aid Report, Studebaker and Bess, Eds. Upper Darby, PA: Monographs in Contemporary Audiology, 87–90 (1982).
[2] B. Rafaely, M. Roccasalva-Firenze, and E. Payne, "Feedback path variability modeling for robust hearing aids," J. Acoust. Soc. Am. **107(5)**, 2665–2673 (2000).
[3] J. Hellgren, T. Lunner, and S. Arlinger, "Variations in the feedback of hearing aids," J. Acoust. Soc. Am. **106(5)**, 2821–2833 (1999).
[4] M. Stinson and G. Daigle, "Effect of handset proximity on hearing aid feedback," J. Acoust. Soc. Am. **115(3),** 1147–1156 (2003).
[5] D. P. Egolf, B. T. Haley, K. A. Weaver, and D. S. Barker, "The hearing aid feedback path: Mathematical simulations and experimental verification," J. Acoust. Soc. Am. **78(5)**, 1578–1587 (1985).
[6] D. P. Egolf, B. T. Haley, H. C. Howell, and S. Legowski, "Simulating the open-loop transfer function as a means for understanding acoustic feedback in hearing aids," J. Acoust. Soc. Am. **85(1)**, 454–467 (1989).
[7] K. Nakao, R. Nishimura, and Y. Suzuki, "Calculation of transfer function of acoustic feedback path for in-the-ear hearing aids with a correction for specific acoustic impedance of a tubule," Acoustical Science and Technology **27(4)**, 242–244 (2006).
[8] J. Maxwell and P. Zurek, "Reducing acoustic feedback in hearing aids," IEEE Transactions on speech and audio processing **3(4)**, 304–313 (1995).
[9] J. M. Kates, "Feedback cancellation apparatus and methods," US Patent, 6072884, 2000.
[10] Y. Haneda, S. Makino and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," IEEE Transactions on speech and audio processing **2(2)**, 320–328 (1994).
[11] Y. Haneda, S. Makino, Y. Kaneda and N. Kitawaki, "Common-acoustical-pole and zero modeling of head-related transfer functions," IEEE Transactions on speech and audio processing **7(2)**, 188–196 (1999).
[12] C. Masterson, S. Adams, G. Kearney and F. Boland, "A method for head related impulse response simlification," in *2009 European Signal Processing Conference*, 2569–2573 (2009).
[13] L. Ljung, and T. Söderström *Theory and practice of recursive identification* (Cambridge, MA: MIT Press) (1983).
[14] P. Chin, R. M. Corless and G. F. Corliss, "Optimization strategies for the approximate GCD problem," in *Proceedings of the 1998 International Symposium on Symbolic and Algebraic Computation*, 228–235 (1998).
[15] J. M. Kates, "Room reverberation effects in hearing aid feedback cancellation," Journal of the Acoustical Society of America **109(1),** 367–378 (2001).
[16] C. M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, USA) (1996).
[17] A. Spriet, *Adaptive Filtering Techniques for Noise Reduction and Acoustic Feedback Cancellation in Hearing Aids* (Ph.D. dissertation, Faculty Eng., Katholiecke Univ. Leuven, Leuven, Belgium) (2004).
[18] G. Ma, F. Gran, F. Jacobsen and F. Agerkvist, "Using a reflection model for modeling the dynamic feedback path of digital hearing aids," J. Acoust. Soc. Am. **127(3)**, 1458–1468 (2010).
[19] J. E. Greenberg, P. M. Zurek and M. Brantley, "Evaluation of feedback-reduction algorithms for hearing aids," J. Acoust. Soc. Am. **108(5)**, 2366–2376 (2000).
[20] A. Spriet, K. Eneman, M. Moonen and J. Wouters, "Objective measures for real-time evaluation of adaptive feedback cancellation algorithms in hearing aids," in *Proceedings of the 16th European Signal Processing Conference*, (2008).

Paper IV

Noise Injection for Feedback Cancellation with Linear Prediction

G. Ma, F. Gran, F. Jacobsen and F. T. Agerkvist

Accepted as a Contributed Paper by the International Congress on Acoustics

Sydney, 2010

Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Noise injection for feedback cancellation with linear prediction

Guilin Ma (1,2), Fredrik Gran (1), Finn Jacobsen (2) and Finn T. Agerkvist (2)

(1) Concept Research, GN ReSound A/S, Lautrupbjerg 9, 2750 Ballerup, Denmark

(2) Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Denmark

PACS: 43.66.Ts,43.60.Ac,43.66.Dc

ABSTRACT

Feedback oscillation is one of the major issues with hearing aids. An efficient way of feedback suppression is feedback cancellation, which uses an adaptive filter to estimate the feedback path. However, the feedback canceller suffers from the bias problem in the feedback path estimate. The recent progress suggests a feedback canceller with linear prediction of the desired signal in order to eliminate the bias when certain conditions are met. However, the bias still remains in many situations, for example when the input signal is voiced speech. Noise injection is investigated in this paper to help reduce the bias further and improve the system performance. Two nearly inaudible noises are proposed: a masking noise, which is tailored to the hearing-aid application, and a linear prediction based noise, which is especially efficient for feedback cancellation with linear prediction. Simulation results show that noise injection can further reduce the feedback estimation error by 1-4 dB and/or increase the stable gain by 3-4 dB, depending on the characteristics of the input signal.

INTRODUCTION

Feedback oscillation is one of the major problems with hearing aids. It limits the maximum gain that can be achieved. A widely adopted approach to feedback suppression is feedback cancellation, where an adaptive filter is used to model the feedback path. However, the closed-loop plant used in continuous-time feedback cancellation systems for hearing aids results in biased estimations of the feedback path when the input and output signals are correlated [1].

Several approaches have been proposed to reduce the bias. Classical approaches include introducing signal de-correlating operations in the forward path or the cancellation path (such as delays or nonlinearities), adding a probe signal to the receiver input, and controlling the adaptation of the feedback canceller , e.g., by means of constrained or bandlimited adaptation [2]. New solutions based on closed-loop identification theory have been investigated in the recent years [3] [4]. As a result, a feedback cancellation system with linear prediction was proposed in [5], which eliminates the bias when certain conditions are met. A combination with the classical solutions were also mentioned in [5], which showed that linear prediction combined with noise injection could improve the system performance further. However, the injected noise used in that work was an audible speech-shaped noise and therefore compromised the sound quality.

This paper proposes two kinds of nearly inaudible noises for injection in the hearing-aid output when the feedback canceller with linear prediction is used. The results show that the injected noise maintains the sound quality, reduces the bias further and increases the stable gain.

The outline of the paper is as follows: in Section , the feedback cancellation system with linear prediction is explained. In Section the adaptation structure for noise injection is discussed and the generation of inaudible noises is described. In Section , simulation results are presented. Concluding remarks are given in Section .

FEEDBACK CANCELLATION WITH LINEAR PREDICTION

Two classes of adaptive procedures for identifying the desired signal model and the feedback path were derived in [5]: a two-channel identification method and a prediction error method (PEM). The latter is found to be preferable for highly non-stationary sound signals and is therefore chosen as the method investigated in this paper.

The diagram of the prediction error model based adaptive feedback canceller (PEM-AFC) used in this work is depicted in FIG. 1. The signal processing path of the hearing aid (the so-called forward path) is denoted by $G(q)$; the acoustic feedback path is denoted by $F(q)$. The receiver and microphone signals are $u[k]$ and $y[k]$, respectively. The filter $\hat{F}_0(q)$ in the feedback cancellation path is an initial estimate of $F(q)$. It is continuously replaced during adaptation by the estimate $\hat{F}(q)$. The external input $x[k]$ is assumed to be an Autoregressive (AR) random process generated from the white noise sequence $w[k]$ and the AR model $H(q)$, which is also referred to as signal model. The FIR filter $A(q)$ is the prediction error filter of the forward-path output $g[k]$. In this diagram, the linear prediction is placed at the receiver end before noise injection which differs from the diagram in [5], where the linear prediction is located at the microphone side and operates on the error signal $e[k]$. This change of placement does not affect the steady-state performance but gives better transient convergence [6]. The prediction error (PE) of $e[k]$ and $u[k]$ are denoted by $e_p[k]$ and $u_p[k]$ respectively. The probe signal (injected noise) is $r[k]$.

Assuming that the input signal $x[k]$ is an AR random process, it has been shown in [5] that when the delay in the forward path is longer than the order of the signal model $H(q)$ or when the probe signal $r[k]$ is introduced, the minimization of $e_p[k]$ leads

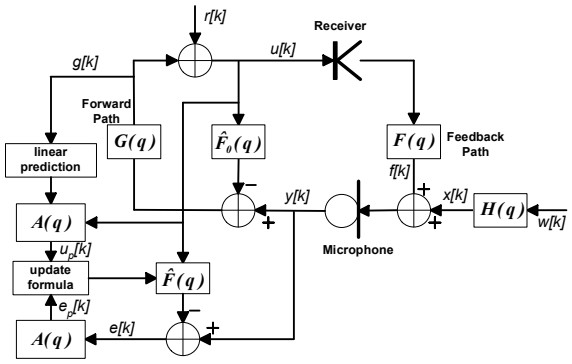


Figure 1: Diagram of the feedback cancellation system with linear prediction (PEM-AFC). Modified from [5].

to:

$$A(q) = H^{-1}(q); \hat{F}(q) = F(q) \tag{1}$$

$$H(q) = \frac{1}{1 + h_1 q^{-1} + \cdots + h_2 q^{-2} + h_{L_H-1} q^{-L_H+1}} \tag{2}$$

where q^{-1} is a discrete-time unit delay operator, $L_H - 1$ is the order of the all-pole filter $H(q)$. $A(q)$, $F(q)$ and $\hat{F}(q)$ are all FIR filters.

Although many signals can be modeled by AR random process, a large set of real-life signals, such as voiced speech and tonal music, can hardly be modeled by a low-order $H(q)$. For these signals, PEM-AFC still suffers from the bias problem due to the remaining spectrally-colored signals[1]. In addition, under-modeling which generally occurs in practice introduces the bias into the estimation. In both cases, a reduction of the remaining bias is necessary.

NOISE INJECTION FOR PEM-AFC

To reduce the bias, the probe signal $r[k]$ can be injected into the signal $g[k]$ before it is sent to the receiver. The injected noise $r[k]$ is generally uncorrelated with the input signal $x[k]$. In Fig. 1, the feedback canceller takes $u[k]$ as the input signal for adaptation. An alternative approach is to use $r[k]$ instead of $u[k]$ in the adaptation of the feedback canceller [7]. However, the low level of the probe signal $r[k]$ (to maintain the sound quality) results in a large excess error or slow adaptation of the adaptive filter $\hat{F}(q)$ [2], which will not be suitable for dynamic situations in the daily life. Therefore only the first approach as shown in Fig. 1 is considered in this paper.

Bias reduction with noise injection

The steady-state analysis of feedback cancellation has revealed that the bias problem is related to the autocorrelation of the input signal $x[k]$ [1]. In PEM-AFC, a similar steady-state analysis can be performed as follows:

$$\hat{\mathbf{f}} = \mathbf{R}^{-1}_{\mathbf{u}_p\mathbf{u}_p} \mathbf{r}_{\mathbf{u}_p y_p}, \tag{3}$$

$$= \mathbf{R}^{-1}_{\mathbf{u}_p\mathbf{u}_p} \mathbf{r}_{\mathbf{u}_p f_p} + \mathbf{R}^{-1}_{\mathbf{u}_p\mathbf{u}_p} \mathbf{r}_{\mathbf{u}_p x_p}, \tag{4}$$

$$\mathbf{R}_{\mathbf{u}_p\mathbf{u}_p} = E\left\{\mathbf{u}_p[k]\mathbf{u}_p^H[k]\right\}, \tag{5}$$

$$\mathbf{r}_{\mathbf{u}_p y_p} = E\left\{\mathbf{u}_p[k] y_p[k]\right\}, \tag{6}$$

$$\mathbf{u}_p[k] = \left[u_p[k], u_p[k-1], \cdots, u_p[k-L+1]\right]^T, \tag{7}$$

$$f_p[k] = A(q)F(q)u[k], \tag{8}$$

$$\hat{F}(q) = \hat{f}_0 + \hat{f}_1 q^{-1} + \cdots + \hat{f}_{L-1} q^{-L+1}, \tag{9}$$

[1]Spectral coloring refers to the fact that certain spectral components are stronger than other spectral components. A spectrally colored signal may be a broad-band (e.g., a speech signal) as well as a narrow-band signal (e.g., a sinusoid or alarm signal) [2].

$$\hat{\mathbf{f}} = \left[\hat{f}_0, \hat{f}_1, \cdots, \hat{f}_{L-1}\right]^T, \tag{10}$$

where $\hat{\mathbf{f}}$ is the coefficient vector of $\hat{F}(q)$, L is its length, the symbol with subscription p denotes the PE of the corresponding signal. The symbol E denotes the expectation operator, and $\mathbf{r}_{\mathbf{u}_p f_p}$ and $\mathbf{r}_{\mathbf{u}_p x_p}$ are defined similarly to equation (6). Suppose that the filter length L is sufficient. The first term in equation (4) approximates the true feedback path, whereas the second term, which represents the correlation between the PE of the desired input signal $x[k]$ and the PE of the processed hearing-aid signal $u[k]$, introduces a bias into the estimate. When the signal $u[k]$ includes the injected noise $r[k]$, the bias term can be reduced. The effectiveness of bias reduction depends on the prediction-error strength of the injected noise $r[k]$ relative to the prediction-error strength of the signal $g[k]$.

However, a loud noise will degrade the sound quality. The injection of noise therefore involves a trade-off between bias reduction and sound quality. To maintain the sound quality, two types of nearly inaudible noises are investigated in the following sections, namely the masking noise and the linear prediction based noise (LP noise).

Masking noise

The Moving-Pictures-Expert-Group (MPEG) standard for audio compression [8] utilizes frequency masking to reduce the number of bits in the intervening transmission. It assumes an audio signal with a sampling frequency of 44.1 kHz and calculates the masking threshold in each sub-band based on the instantaneous spectrum of the signal by using a 512-point FFT.

To make it possible and affordable for hearing-aid application, the calculation is modified to operate with the sampling frequency of 16 kHz and 64-point FFT. The loss in sound quality due to such a modification is found to be insignificant. In addition, since feedback whistling usually occurs below 6 kHz, the masking threshold is not calculated for those sub-bands above 6 kHz.

The masking noise is generated by shaping the white noise sequence with the calculated masking threshold so that it has the same average spectrum as the masking threshold. It should be stressed that the masking noise formed in this way is still audible because the psychoacoustic model used to calculate the masking thresholds is established for sinusoids instead of noise. Therefore the masking noise generated from the masking thresholds is attenuated by 14 dB in this work to make it nearly inaudible.

LP noise

In order to avoid the degradation of sound quality while providing as strong a force as possible to reduce the bias, the injected noise $r[k]$ should be strong in those frequency bands where the desired signal is strong and weak where it is weak. LP noise is proposed to achieve this idea efficiently.

Instead of calculating the spectrum of the outgoing signal $g[k]$, LP noise is generated by filtering a white noise sequence with the inverse of the prediction-error filter $A(q)$ as shown in the equations below:

$$g_p[k] = A(q)g[k] \tag{11}$$

$$p[k] = \frac{1-\beta}{1-\beta q^{-1}} g_p[k]^2 \tag{12}$$

$$r[k] = \alpha\sqrt{p[k]}A^{-1}(q)n[k] \tag{13}$$

where $p[k]$ estimates the smoothed prediction-error power of $g[k]$ by passing the instant power through a low-pass IIR filter configured by a smoothing factor β, which is set to 0.996 in

this paper. The white noise sequence $n[k]$ has unit variance. The parameter α controls the strength of the LP noise. $\alpha = 0.3$ is found to be low enough to produce nearly inaudible noises for most signals.

Noise generated in this way has two advantages: first, the computation load is very low as $A(q)$ is ready to use after the linear prediction stage; secondly, when $r[k]$, embedded in $u[k]$, goes through $A(q)$ to form the prediction error signal $u_p[k]$, it becomes the white noise sequence $n[k]$ again, which has ideal autocorrelation properties. To ensure the stability of the IIR filter $A^{-1}(q)$, the linear prediction has to yield a stable model.

SIMULATION RESULTS

To evaluate the performance of PEM-AFC with noise injection, the system in the Fig. 1 is simulated with a sampling frequency of 16 kHz. The feedback path is 50-order, measured from a commercial behind-the-ear (BTE) device and normalized so that the maximum stable gain is 0 dB without feedback cancellation, as illustrated in Fig. 2. The feedback model $\hat{F}(q)$ is also 50-order and the linear prediction-error filter $A(q)$ is 21-order. Levinson-Durbin algorithm, which yields stable models, is used to linear predict $g[k]$ with a frame length of 168 samples. The forward path $G(q)$ consists of a delay of 24 samples and a linear gain. The adaptation algorithm for feedback canceller is the recursive-least-square (RLS) algorithm with a forgetting factor $\lambda = 1$. Three types of input signals are investigated: a 3 kHz pure tone in white noise (to simulate the background noise and the microphone noise) whose power is -40 dB below the tone, male speech signal and guitar music signal. All the signals are 12-second long.

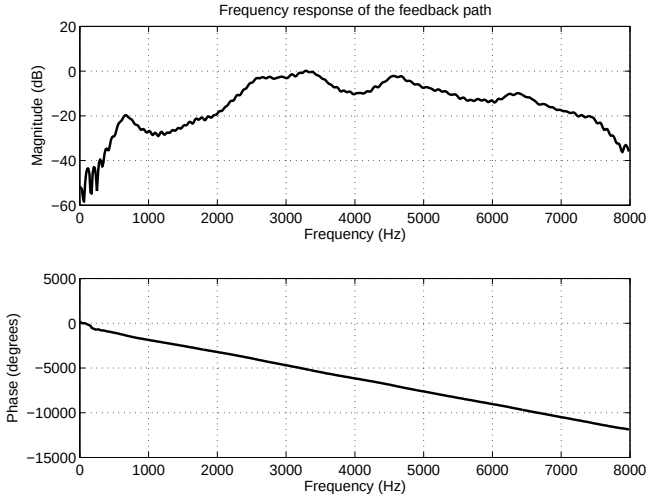


Figure 2: The frequency response of the feedback path of 50 orders based on the measurement of a commercial BTE hearing aid.

As mentioned in Section , in PEM-AFC, the amount of bias reduction actually depends on the relative strength of the prediction error of the injected noise. To illustrate the effectiveness of the two inject noises, it is first assumed that the feedback cancellation is perfect, i.e., $\hat{F}(q) = F(q)$. When the 3 kHz sinusoid in white noise is input into the system, the power spectral density (PSD) of the signal $g[k]$, the injected LP noise, the injected masking noise and PSD of their corresponding prediction errors are plotted in Fig. 3.

In the figure, it can be seen that the PSD of the prediction error of the signal $g[k]$, i.e. $g_p[k]$, exhibits a peak at 3 kHz, which indicates that $g_p[k]$ is still a highly autocorrelated signal and therefore the bias remains in the system after linear prediction. The PSD of the injected LP noise resembles the PSD of the signal $g[k]$, and the prediction error of the LP noise is white as expected. In the 3-kHz frequency band, the prediction error of the LP noise is as strong as the peak in the PSD of $g_p[k]$. Hence, the LP noise can provide a significant force to reduce the bias. In those frequency bands around 3 kHz (from 2.72 kHz to 2.97 kHz and from 3.03 kHz to 3.28 kHz), the prediction error of the masking noise has almost the same power as that of $g_p[k]$. Therefore, the masking noise should help more in bias reduction at those frequencies than the LP noise.

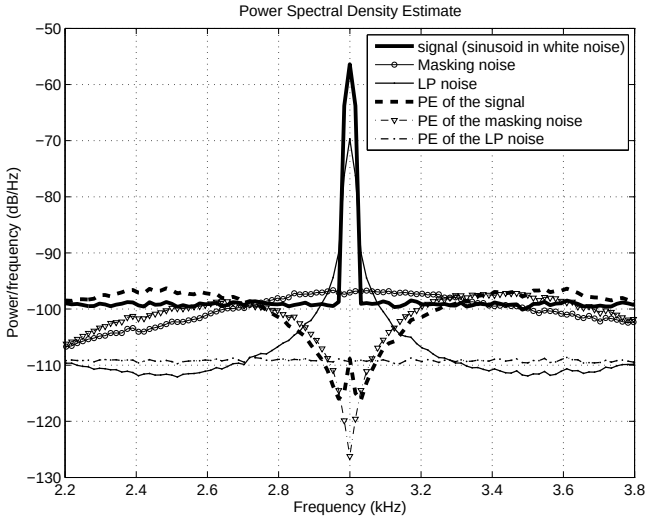


Figure 3: Power spectral density of the signal (a 3-kHz sinusoid in white noise), the injected masking noise, the injected LP noise and their corresponding prediction errors.

To assess the performance quantitatively, we used signal to feedback residual error (SFRR) as the measure:

$$SFRR = 10\log_{10}\frac{\sum_{k=1}^{N}(F(q)u[k]-\hat{F}_0(q)u[k])^2}{\sum_{k=1}^{N}(x[k])^2} \qquad (14)$$

where N is the total number of samples used in the simulation. $x[k]$ and $u[k]$ are both zero-mean. SFRR represents the ratio of the input signal to the distortion resulted from the feedback estimation error. This measure is more proper for narrow-band signals than the traditional broad-band measure maximum stable gain (MSG) because stable gains outside the signal band are not very meaningful.

The simulation results are shown in Fig. 4. For a sinusoid in white noise, the injected LP noise yields a 1-4 dB improvement of SFRR, and the masking noise provides 4 dB extra stable gain (the stable gain is defined as the point where SFRR begins to drop steeply). For music, the LP noise improves SFRR by 1 dB whereas the masking noise improves the SFRR slightly. For speech, the LP noise and the masking noise improve SFRR slightly and provide 3-4 dB extra stable gain.

In general, the injected LP noise performs the best (1-4 dB SFRR increase) before the system becomes instable, whereas the masking noise is better at providing additional stable gain. The reason is that when the system is stable, the LP noise has a stronger force to reduce the bias (c.f. Fig. 3). However, when the whistle is about to occur, the linear prediction in PEM-AFC tends to model the strong feedback residual signal instead of the desired signal. The LP noise generated with these wrong linear prediction coefficients does not help very much in the bias reduction. The masking noise, which on the other hand, takes into account the spectrum of the whole signal including both the desired signal and the feedback residual signal, may still provide forces in keeping the system from instability.

In the simulation, it was also found that the LP noise and masking noise injected in the traditional feedback cancellation system without linear prediction provided very little improvement. This is in agreement with the findings in [5]. Therefore the results are omitted in Fig. 4. Noise injection is more effective in feedback cancellation with linear prediction probably because of two reasons: Firstly, linear prediction reduces the bias and therefore slows down the speed that the system can go wrong at. As input signals usually have weak periods with low amplitude from time to time (due to the amplitude fluctuation and pauses), the noise injection may help the system to go back quickly onto the right track in those periods. Even when the adaptation will still be biased afterwards, the system may keep stable before the next weak period is hit. However, if there is no linear prediction, the efforts noise injection has made during the weak periods will be overruled quickly by the large bias and the whistle may occur immediately. Secondly, linear prediction is performed on noisy data, so over- and under-modeling can both occur, which averages out part of the remaining bias. In other words, the injected noise actually plays a relatively larger role than expected.

CONCLUSIONS

This paper investigates the inaudible noise injection in the feedback cancellation with linear prediction. Two inaudible noises are proposed: the masking noise, which is tailored to the hearing-aid application based on the MPEG standards, and the LP noise, which is proposed specially for the feedback canceller with linear prediction to achieve an efficient implementation. The effect of noise injection is evaluated for tonal signal, speech and music. It is shown that noise injection can reduce the feedback estimation error by 1-4 dB and/or increase the stable gain by 3-4 dB, depending on the characteristics of the input signal.

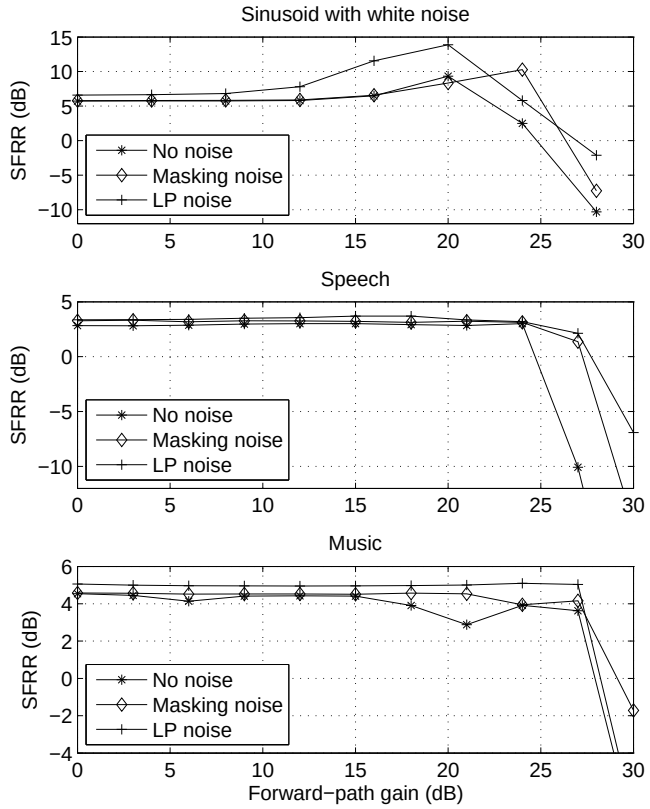


Figure 4: SFRR of PEM-AFC with no noise, LP noise injected and masking noise injected when the input is a sinusoid with noise, male speech and guitar music respectively.

REFERENCES

[1] M. G. Siqueira and A. Alwan. Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids. *IEEE Transactions on speech and audio processing*, 8(4):443–453, 2000.

[2] A. Spriet. *Adaptive Filtering Techniques for Noise Reduction and Acoustic Feedback Cancellation in Hearing Aids*. Ph.D. dissertation, Faculty Eng., Katholiecke Univ. Leuven, Leuven, Belgium, 2004.

[3] J. Hellgren and U. Forssell. Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 9(8):906–913, 2001.

[4] N. A. Shusina and B. Rafaely. Feedback cancellation in hearing aids based on indirect close-loop identification. In *Proc. IEEE Benelux Signal Process. Symp*, pages 177–180, 2002.

[5] A. Spriet, I. Proudler, M. Moonen, and J. Wouters. Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal. *IEEE Transactions on Signal Processing*, 53(10):3749–3763, 2005.

[6] M. M. Yasmin and M. B. Jose Carlos. Mean weight behavior of coupled lms adaptive systems applied to acoustic feedback cancellation in hearing aids. In *Proc. ICISP*, pages 527–535, 2008.

[7] H. R. Skovgaard. Hearing aid compensating for acoustic feedback. In *Patent, US*, page 5680467, 1997.

[8] ISO/IEC. *Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 3: Audio*. ISO/IEC std. 11172-3, 1993.

Paper V

A New Approach for Modelling the Dynamic Feedback Path of Digital Hearing Aids

G. Ma, F. Gran, F. Jacobsen and F. T. Agerkvist

Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing

Presented in Taipei, Taiwan, April 2009

A NEW APPROACH FOR MODELLING THE DYNAMIC FEEDBACK PATH OF DIGITAL HEARING AIDS

Guilin Ma[1,2]*, Fredrik Gran*[1]*, Finn Jacobsen*[2] *and Finn Agerkvist*[2]

[1]GN ReSound A/S, Lautrupbjerg 9, 2750 Ballerup, Denmark.

[2] Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark,
Oersteds Plads, Building 352 DK-2800 Kgs. Lyngby, Denmark.

ABSTRACT

This paper proposes a reflection model for the dynamic feedback path of digital hearing aids and compares it with two existing models: a direct model and an initialization model, based on the measured dynamic feedback paths. The comparison shows that the proposed model is superior to the existing two models in terms of maximum stable gain (MSG). For hearing aids with dual microphones, the possibility of relating the two dynamic feedback paths is also investigated. It is shown that in a complicated acoustic environment, the relation between the two feedback paths can be very intricate and difficult to exploit in modelling the dynamic feedback paths.

***Index Terms*—** Feedback cancellation, hearing aids, dynamic feedback path modelling, delay estimation.

1. INTRODUCTION

Feedback is one of the major problems with hearing aids. It limits the maximum gain that can be achieved. A widely adopted approach to feedback suppression is feedback cancellation, where an adaptive filter is used to model the feedback path. The output of the filter is regarded as the instantaneous estimation of the feedback signal and is subtracted from the input signal to remove the feedback.

The maximum stable gain (MSG) obtained by using a feedback canceller depends on how accurately the feedback path can be estimated. A perfect match between modelled and real feedback path will cancel the feedback signal completely, and the system will be stable for any amount of amplification [1]. In practice, however, the feedback path may be subject to dramatic changes, e.g., when the user picks up the phone. In these adverse situations, the feedback canceller usually has problems in obtaining an accurate estimate of the feedback path due to its slow convergence and/or biased adaptation. A whistle is therefore easily triggered. This has become the major concern of hearing aid users with feedback problems today.

In order to improve the performance of a feedback canceller in dynamic situations, the model of dynamic feedback path should be investigated. However, to our best knowledge, very little research has been carried out in analyzing the dynamic change of feedback path in real life. This paper proposes a reflection model for the dynamic feedback path and compares it with two existing models using data from measurements of dynamic feedback paths. The comparison shows that the proposed model is superior to the existing two models in terms of MSG. For hearing aids with dual microphones, the possibility of relating the two feedback paths is also investigated. It is shown that in a complicated acoustic environment, the relation between the two feedback paths can be very intricate and difficult to exploit in modelling the dynamic feedback paths.

The outline of the paper is as follows: in section 2 two traditional models are explained and a new reflection model is proposed. In section 3 the measurement is described and the results are given based on the measured data. Concluding remarks and directions for future work are given in section 4.

Please address all correspondence to Guilin Ma, *gm@elektro.dtu.dk* Technical University of Denmark.

2. MODELS FOR THE DYNAMIC FEEDBACK PATH

The general diagram of feedback cancellation is depicted in Fig. 1. The idea of feedback cancellation is to adjust the parameters θ in the feedback model so that the modelled feedback path $\hat{b}(n,\theta)$ approximates the true feedback $b(n)$ as close as possible. The output $v(n)$ is the instantaneous estimation of the feedback signal $f(n)$ and is subtracted from the input signal $s(n)$ to remove the feedback.

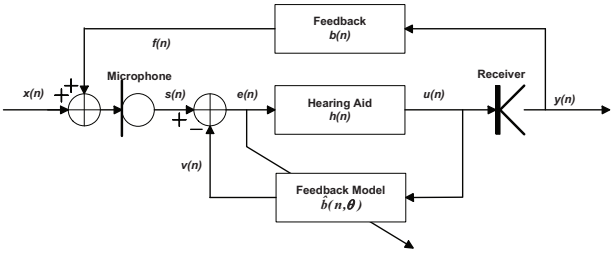


Fig. 1. General diagram of feedback cancellation. The input to the hearing-aid processing is $s(n)$, which is the sum of desired input signal $x(n)$ and the feedback signal $f(n)$. The processed hearing-aid signal is $u(n)$. The signal output into the ear canal is $y(n)$. The impulse response of the feedback path is $b(n)$, and $v(n)$ is the estimation of $f(n)$ from the modelled feedback path $\hat{b}(n,\theta)$.

In real life, the impulse response of the feedback path $b(n)$ is time-varing and can change dramatically. An example is shown in Fig. 2, where the impulse response of the feedback path is measured without any enclosure and with a palm wrapping around the hearing aid, which mimics the situation when the user picks up the phone. As seen from the figure, both the impulse response and the frequency response change remarkably when the hearing aid is enclosed.

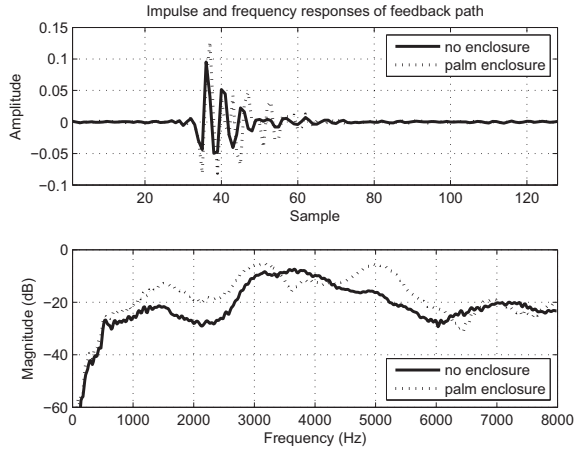


Fig. 2. Impulse responses and frequency responses of feedback paths with (dotted lines) and without (solid lines) palm enclosure.

2.1. Measure of the feedback models

In principle, the impulse response $b(n)$ has an infinite duration. However, as shown in Fig. 2, the amplitude of $b(n)$ decays very fast. Assume that the truncated impulse responses of $b(n)$ and $\hat{b}(n,\theta)$ with length L are sufficient to represent the true feedback path and the feedback model respectively. One natural way of obtaining the optimal parameters θ_{opt} for the feedback model is to minimize the difference between the truncated feedback model and the actual feedback path. This, formulated in the frequency domain, is given by

$$\theta_{opt} = \arg\min_{\theta} \|\mathbf{F}^H(\hat{\mathbf{b}}(\theta) - \mathbf{b})\|_2^2, \quad (1)$$

$$\mathbf{b} = (\ b(0), \quad \dots \quad , b(L-1)\ )^T, \quad (2)$$

$$\hat{\mathbf{b}}(\theta) = (\ \hat{b}(0,\theta), \quad \dots \quad , \hat{b}(L-1,\theta)\ )^T, \quad (3)$$

$$\mathbf{F} = (\ \mathbf{f}_0, \quad \mathbf{f}_1, \quad \dots \quad \mathbf{f}_{L-1}\ ) \quad (4)$$

$$\mathbf{f}_k = (\ 1, \quad e^{j\omega_k}, \quad \dots \quad , e^{j\omega_k(L-1)}\ )^T, \quad (5)$$

where $\omega_k = 2\pi l/L, l = 0, 1, \dots, L-1$, $\mathbf{F}$ is the Fourier matrix, and $(\cdot)^T$ denotes the transpose of $(\cdot)$.

To evaluate the performance of a feedback model, MSG is often used, which is determined by the frequency at which the mismatch between the feedback model and the actual feedback path is the largest [2]. We assume that in all the circumstances, the parameters in the feedback model θ can converge to the optimal solution θ_{opt} fast and accurately enough[1]. The MSG of the model is therefore the MSG with converged parameters, denoted as MSG_c,

$$\text{MSG}_c = 20\log_{10}\left(\min_k \frac{1}{|\mathbf{f}_k^H(\hat{\mathbf{b}}(\theta_{opt}) - \mathbf{b})|}\right). \quad (6)$$

With a specific model and parameters θ, MSG_c is the highest achievable MSG, which is in fact limited by the amount of under-modelled feedback path, the residual feedback path that cannot be modelled due to the limited degrees of freedom in the parameter θ and/or the lack of flexibility in the model form. A more descriptive model with larger degrees of freedom in the parameters θ will yield less under-modelling and larger MSG_c.

In the following text, three models will be described and MSG_c will be computed to evaluate and compare these models.

[1]The feedback canceller usually suffers from the problem of slow convergence and biased adaptation. These two topics, however, are irrelevant with the model of feedback path. Therefore they are not considered in this paper to simplify the model comparison.

2.2. Direct model

One typical form of feedback model is composed of a pre-filtering and an adaptive filter, which is usually FIR (Finite Impulse Response) since IIR (Infinite Impulse Response) adaptive filtering suffers from the problem of instability and local minima [3]. Let $b_0(n)$ and $w(n)$ denote the impulse response of the pre-filtering and the adaptive filter respectively. The feedback model is the convolution of $b_0(n)$ and $w(n)$,

$$\hat{b}(n,\theta) = w(n) \odot b_0(n) = \sum_{l=0}^{M-1} w(l)b_0(n-l), \quad (7)$$

$$\theta = \{w(n), b_0(n)\}, \quad (8)$$

where M is the order of $w(n)$, and $\odot$ is the convolution operator.

In the "direct model", the pre-filtering is simply a delay of D samples:

$$b_0(n) = \begin{cases} 1, n = D+1 \\ 0, otherwise \end{cases}. \quad (9)$$

Since $b(n)$ usually starts with a certain physical delay (see Fig. 2), the introduction of a corresponding delay D renders a better use of the limited number of taps in the adaptive filter $w(n)$. To calculate MSG_c, the optimal parameters, i.e., $w_{opt}(n)$ and D_{opt} should be obtained first by solving equations (1)-(5), (7)-(8) and (9).

This is a nonlinear optimization problem. However, it can be solved easily in numerical ways. As a special case, if the delay D is fixed, it reduces to a simple optimization problem with the following solution

$$\mathbf{w}_{opt} = (\ b(D+1), \quad \dots \quad , b(D+M)\ )^T. \quad (10)$$

With the optimal parameters, MSG_c can be calculated from equation (6).

2.3. Initialization model

To model the feedback path accurately, the direct model in 2.2 usually needs a very high-order adaptive FIR filter $w(n)$ to cover the "active" range in Fig. 2. One way to reduce the number of orders needed for modelling the dynamic feedback path is to use an initialization as proposed in [4], which is a measurement of feedback path in a static situation without any reflectors or enclosures near the hearing aid. When the hearing aid is put into use in daily life, to capture the time-varying dynamic feedback path, the adaptive filter $w(n)$ only needs to model the part that is different from the static initialization. Since the impulse responses of microphone, receiver, etc. will not change from the static initialization to the dynamic situation, this different part can be modelled by an adaptive filter with a lower order.

Let $b_0(n)$ denote the impulse response of the static feedback path obtained in the initialization. The modelled feedback model is the same as in equation (7) with

$$\theta = \{w(n)\}, \tag{11}$$

The impulse response $b_0(n)$ is truncated to $L - M + 1$ samples here so that the length of the convolution between $w(n)$ and $b_0(n)$ equals L. In practice $b_0(n)$ can be implemented by an IIR filter [4].

When $w(n)$ is real-valued, the optimal parameter for the initialization model $w_{opt}(n)$ can be found by solving a linear least square problem with equation (1)-(5), (7) and (11):

$$\begin{aligned}
\mathbf{w}_{opt} &= (diag(\mathbf{F}^H\tilde{\mathbf{b}}_0)\tilde{\mathbf{F}}^H)^+(\mathbf{F}^H\mathbf{b}), && (12)\\
\tilde{\mathbf{b}}_0 &= (\ \mathbf{b}_0^T,\quad \mathbf{0}_{1\times(M-1)}\ )^T, && (13)\\
\mathbf{b}_0 &= (\ b_0(0),\quad \dots\quad , b_0(L-M)\ )^T, && (14)\\
\tilde{\mathbf{F}} &= \begin{pmatrix} f_0(0) & \dots & f_{L-1}(0) \\ \cdots & \cdots & \cdots \\ f_0(M-1) & \dots & f_{L-1}(M-1) \end{pmatrix} && (15)
\end{aligned}$$

where $diag(\cdot)$ is a diagonal matrix with diagonal elements $(\cdot)$, $(\cdot)^+$ is a pseudo-inverse defined as $(\cdot)^+ = ((\cdot)^T(\cdot))^{-1}(\cdot)^T$, and $\mathbf{0}_{1\times(M-1)}$ represents a zero vector of length $M-1$.

The optimal solution represents the adaptive filter $w(n)$ of length M that produces MSG_c when being concatenated with the initialization filter $b_0(n)$ to model the dynamic feedback path $b(n)$.

2.4. Reflection model

In section (2.3), the initialization model formulated in equation (7), can also be regarded as a weighted sum of the initialization $b_0(n)$ and its delayed replicas with integer delays. We generalize it to a new model with fractional delays, i.e.,

$$\begin{aligned}
\hat{b}(n,\theta) &= \sum_{l=0}^{M-1} w(l)b_0(n-d_l), && (16)\\
\theta &= \{w(l), d_l\}, && (17)
\end{aligned}$$

where d_l is the delay of the l-th replica, $d_l > d_{l-1} \geq 0, l = 1, \cdots, M-1$.

These delayed replicas can be interpreted as physical reflections with delay d_l and gain $w(l)$. This model is thus named "reflection model". Compared with the initialization model, the reflection model is more precise because it mimics what happens in the physical world. For example, when the user picks up the phone, the feedback path consists of a direct path and multiple reflections with possibly fractional delays. The direct path can be approximated by the initialization since it is done in the static situation without any reflectors or enclosures near the hearing aid. In fact, when $d_l = l$, the reflection model is identical to the initialization model. Therefore, the reflection model is more general and expected to capture the dynamic feedback path better than the initialization model.

The optimal delays $d_{l,opt}$ and weights $w_{opt}(l)$ for the reflection model can be found by solving the optimization problem (1)-(5), (16) and (17), which is a nonlinear optimization problem. However, efficient time delay estimation techniques exist to address the problem. An iterative search of $w_{opt}(l)$ and $d_{l,opt}$ proposed in [5] is found to be very robust. We first cross-correlate $b(n)$ and $b_0(n)$ in the frequency domain to find the coarse delays and gains of the replicas by identifying the peaks of the cross-correlation. Later an iterative search is performed by keeping one replica of $b_0(n)$ at a time (removing the other identified replicas from $b(n)$), repeating the cross correlation and locating the peak to find a better delay and gain estimation for that replica. This process is iterated until the relative change of the cost function in equation (1) is below the threshold.

2.5. Models for dual-microphone hearing aids

For hearing aids with dual microphones, the feedback problem involves two feedback paths, denoted as $b_1(n)$ and $b_2(n)$. One way to deal with the two paths is to model them individually by using one of the three models described above. An alternative approach is to fit one feedback path with the other path. There are two ways of fitting, similar to the initialization model and reflection model respectively.

The first approach for the fitting is:

$$\hat{b}_1(n) = \sum_{l=0}^{M-1} w(l)b_2(n-l), \tag{18}$$

The second is:

$$\hat{b}_1(n) = \sum_{l=0}^{M-1} w(l)b_2(n-d_l), \tag{19}$$

The optimal weights $w_{opt}(n)$ and/or delays $d_{l,opt}$ can be found in similar ways described in section 2.3 and 2.4.

3. MEASUREMENT AND RESULTS

The static and dynamic feedback paths are measured using a commercial open-fitting behind-the-ear (BTE) device from GN ReSound A/S. For each feedback path, MSG_c of the three models and the models for dual-microphone hearing aids is calculated by optimizing the parameters in the model.

3.1. Measurements

The hearing aid is mounted on the head of Kemar Manikin Type 45BA made by G.R.A.S Sound & Vibration A/S. The impulse response of the feedback path is measured by sending out a maximum-length sequence (MLS) with a period of 255 samples through the receiver. One thousand periods are repeated to obtain a high SNR for the feedback path response relative to random room noise. The sampling frequency is 16 kHz. The detailed procedure of the impulse response measurement can be found in [2].

The measurement included two steps: First an initialization was carried out to measure the static feedback path without any reflectors or enclosures. Then dynamic feedback paths were created to mimic the most adverse situations for feedback cancellation in real life by a special setup: A rigid surface facing the hearing aid was moved along the lateral side gradually towards the hearing aid and outwards later. The perpendicular distance between the rigid surface and the hearing aid was kept at around three centimetres during the movement. The impulse responses were measured and five representative snapshots were selected for analysis. Two additional dynamic feedback paths were measured with a open palm facing the hearing aid on its lateral side at a distance of three centimetres and with a palm wrapping around the hearing aid. Altogether eight impulse responses were measured including one initialization (static feedback) and seven dynamic feedback paths.

3.2. Results

For each measured dynamic feedback path, the parameters in the models were first optimized to calculate the MSG_c. The filter length M was varied from 1 to 50. In the direct model, the delay D is not fixed to achieve the best performance.

It is found that for all the seven dynamic paths and all the values M, the reflection model outperforms the initialization model and the direct model in terms of MSG_c. The direct model performs the worst in almost all the cases. To demonstrate the performance of each model in dynamic situations, MSG_c is averaged over the seven dynamic paths. The results are illustrated in Fig. 3 and Fig. 4. The results for the dual-microphone models, denoted as "2 channel", are also included.

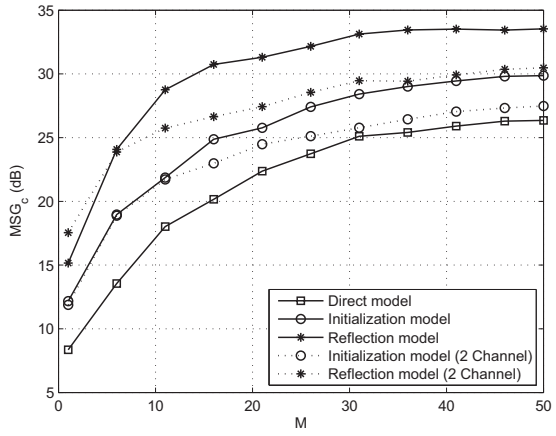


Fig. 3. Comparison of the models for dynamic feedback path modelling.

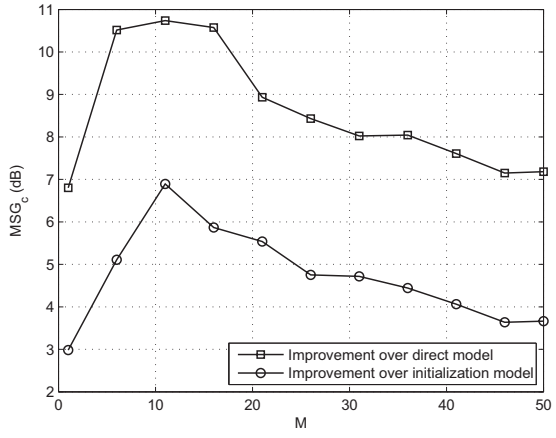


Fig. 4. The MSG_c improvement of the reflection model over direct model and initialization mode for dynamic feedback path modelling.

As seen from the figures, the reflection model is superior to the other two models especially when M is around 11. In practice, M is usually chosen between 10 to 20 to assure a fast convergence. In this region, the reflection model yields 5-7 dB higher MSG_c than the initialization model and 9-11 dB higher MSG_c than the direct model. To achieve a 25dB MSG_c, the direct model needs 31 orders and the initialization model needs 16 orders, whereas, the reflection model only needs 7 replicas of the initialized impulse response.

It is also noted that the dual-microphone models by relating the two feedback paths do not give any benefit. This is because in a complex acoustic environment, the relation between the two feedback paths can be very complicated and even more difficult to model than the feedback paths themselves.

4. CONCLUSIONS AND FUTURE WORK

This paper proposes a novel reflection model for the dynamic feedback path in digital hearing aids. The results based on the measurement of a commercial hearing aid show that the proposed model has better ability in capturing the dynamic feedback path and is superior to the existing two models in terms of MSG.

The results also give the minimum order of the adaptive filter in the two existing models to achieve a certain MSG in the dynamic situations, which could serve as a useful indication in practice for choosing the order of the adaptive filter in the feedback canceller.

Moreover, this paper investigates the possibility of relating the two feedback paths of a dual-microphone hearing aid for modelling the dynamic feedback paths. It is shown that in a complex acoustic environment, the relation between the two feedback paths can be very complicated and difficult to exploit to yield better models.

The drawback of the proposed method is the complexity in estimating the fractional delays. The future work is to find an efficient way of estimating the delays and investigate how to use this reflection model in an on-line adaptation to improve the performance of feedback cancellation in dynamic situations.

5. REFERENCES

[1] J.M. Kates, *Digital hearing aids*, Plural Publishing, 2008.

[2] J.M. Kates, "Room reverberation effects in hearing aid feedback cancellation," *Journal of the Acoustical Society of America*, vol. 109(1), pp. 367–378, 2001.

[3] J.A. Maxwell and P.M. Zurek, "Reducing acoustic feedback in hearing aids," *IEEE Transactions on speech and audio processing*, vol. 3(4), pp. 304–313, 1995.

[4] J.M. Kates, "Constrained adaptation for feedback cancellation in hearing aids," *Journal of the Acoustical Society of America*, vol. 106(2), pp. 1010–1019, 1999.

[5] Jian Li and Renbiao Wu, "An efficient algorithm for time delay estimation," *IEEE Transactions on Signal Processing*, vol. 46(8), pp. 2231–2235, 1998.

Bibliography

[1] A. Spriet. *Adaptive Filtering Techniques for Noise Reduction and Acoustic Feedback Cancellation in Hearing Aids*. Ph.D. dissertation, Faculty Eng., Katholiecke Univ. Leuven, Leuven, Belgium, 2004.

[2] H. Dillon. *Hearing Aids*. Thieme Medical Publishers, Inc., 2001.

[3] S. F. Lybarger. Acoustic feedback control. *The Vanderbilt Hearing-Aid Report, Studebaker and Bess, Eds. Upper Darby, PA: Monographs in Contemporary Audiology*, pages 87–90, 1982.

[4] S. Kochkin. Marketrak viii: Consumer satisfaction with hearing aids is slowly increasing. *The Hearing Journal*, 63(1):19–32, 2010.

[5] J. M. Kates. *Digital Hearing Aids*. Plural Publishing, 2008.

[6] A. Schaub. *Digital Hearing Aids*. Thieme Medical Publishers, Inc., 2008.

[7] A. Vonlanthen and H. Arndt. *Hearing Instrument Technology for the Hearing Health Care Professional (3rd Edition)*. Thomson Delmar Learning, Clifton Park, NY, 2006.

[8] M. Valente. *Hearing Aids: Standards, Options, and Limitations (2nd Edition)*. Thieme Medical Publishers, Inc., 2002.

[9] S. Kochkin. Customer satisfication with hearing instruments in the digital age. *The Hearing Review*, 58(9):30–43, 2005.

[10] S. Kochkin. Marketrak vi: Consumers rate improvements sought in hearing instruments. *The Hearing Review*, 9(11):18–22, 2002.

[11] B. Rafaely, M. Roccasalva-Firenze, and E. Payne. Feedback path variability modeling for robust hearing aids. *Journal of the Acoustical Society of America*, 107(5):2665–2673, 2000.

[12] J. Benesty and Y. Huang. *Adaptive Signal Processing: Applications to Real-World Problems*. Springer, 2003.

[13] J. Hellgren, T. Lunner, and S. Arlinger. Variations in the feedback of hearing aids. *Journal of the Acoustical Society of America*, 106(5):2821–2833, 1999.

[14] M. R. Stinson and G. A. Daigle. Effect of handset proximity on hearing aid feedback. *Journal of the Acoustical Society of America*, 115(3):1147–1156, 2003.

[15] D. P. Egolf, B. T. Haley, K. A. Weaver, and D. S. Barker. The hearing aid feedback path: Mathematical simulations and experimental verification. *Journal of the Acoustical Society of America*, 78(5):1578–1587, 1985.

[16] D. P. Egolf, B. T. Haley, H. C. Howell, and S. Legowski. Simulating the open-loop transfer function as a means for understanding acoustic feedback in hearing aids. *Journal of the Acoustical Society of America*, 85(1):454–467, 1989.

[17] K. Nakao, R. Nishimura, and Y. Suzuki. Calculation of transfer function of acoustic feedback path for in-the-ear hearing aids with a correction for specific acoustic impedance of a tubule. *Acoustical Science and Technology*, 27(4):242–244, 2006.

[18] J. M. Kates. Feedback cancellation in hearing aids: a computer simulation results. *IEEE Transactions on Signal Processing*, 30(3):553–562, 1991.

[19] A. Spriet, I. Proudler, M. Moonen, and J. Wouters. Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal. *IEEE Transactions on Signal Processing*, 53(10):3749–3763, 2005.

[20] J. M. Kates. *Feedback cancellation in a hearing aid with reduced sensitivity to low-frequency tonal inputs*. US Patent, US 6831986, 2004.

[21] J. Hellgren. Analysis of feedback cancellation in hearing aids with filtered-x lms and the direct method of closed loop identification. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 10(2):119–131, 2002.

[22] J. A. Maxwell and P. M. Zurek. Reducing acoustic feedback in hearing aids. *IEEE Transactions on speech and audio processing*, 3(4):304–313, 1995.

[23] J. M. Kates. Constrained adaptation for feedback cancellation in hearing aids. *Journal of the Acoustical Society of America*, 106(2):1010–1019, 1999.

[24] J. L. Nielsen and U. P. Svensson. Performance of some linear time-varying systems in control of acoustic feedback. *Journal of the Acoustical Society of America*, 1(106):240–254, 1999.

[25] M. R. Schroeder. Improvement of acoustic-feedback stability by frequency shifting. *Journal of the Acoustical Society of America*, 36(9):1718–1724, 1976.

[26] D. K. Bustamante, T. L. Worrall, and M. J. Williamson. Measurement and adaptive suppression of acoustic feedback in hearing aids. In *Proceedings of 1989 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2017–2020, 1989.

[27] S. Haykin. *Adaptive Filter Theory, 4th edition*. Prentice-Hall, 2002.

[28] M. G. Siqueira and A. Alwan. Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids. *IEEE Transactions on speech and audio processing*, 8(4):443–453, 2000.

[29] M. A. Stone and B. C. J. Moore. Tolerable hearing aid delays. i. estimation of limits imposed by the auditory path alone using simulated hearing losses. *Ear and Hearing*, 20(3):182–191, 1999.

[30] H. A. L. Joson, F. Asano, Y. Suzuki, and S. Toshio. Adaptive feedback cancellation with frequency compression for hearing aids. *Journal of the Acoustical Society of America*, 94(6):3248¨C3254, 1993.

[31] C. Boukis, D. P. Mandic, and A. G. Constantinides. Toward bias minimization in acoustic feedback cancellation systems. *Journal of the Acoustical Society of America*, 121(3):1529–1237, 2007.

[32] H. R. Skovgaard. *Hearing aid compensating for acoustic feedback*. US Patent, US 5680467, 1997.

[33] H. F. Chi, S. X. Gao, S. D. Soli, and A. Alwan. Band-limited feedback cancellation with a modified filtered-x lms algorithm for hearing aids. *Speech Communication*, 39(1):147–161, 2003.

[34] A. Spriet, G. Rombouts, M. Moonen, and J. Wouters. Adaptive feedback cancellation in hearing aids. *Journal of the Franklin Institute*, 343(6):545–573, 2006.

[35] S. Lee, I. Kim, and Y. Park. An efficient feedback canceler for hearing aids based on approximated affine projection. *Lecture Notes in Computer Science*, 4115:711–720, 2006.

[36] J. M. Kates and J. L. Melanson. *Feedback cancellation apparatus and methods utilizing adaptive reference filter mechanisms*. US Patent, US 6434247, 2002.

APPENDIX

A

Additional Publication

This appendix includes one additional paper based on part of my Master's thesis on source separation approaches for feedback cancellation. This paper was presented at the IEEE International Symposium on Signal Processing and Information Technology 2007 in Cairo, Egypt and is not covered in the dissertation.

Paper VI

Monaural Separation of Dependent Audio Sources Based on a Generalized Wiener Filter

G. Ma, F. T. Agerkvist and J. B. Luther

Proceedings of the IEEE International Symposium on Signal Processing and Information Technology

Presented in Cairo, Egypt, December 2007

2007 IEEE International Symposium
on Signal Processing and Information Technology

Monaural Separation of Dependent Audio Sources Based on a Generalized Wiener Filter

Guilin Ma[1], Finn T. Agerkvist[1], Jim Benjamin Luther[2]
[1] Acoustic Technology, Ørsted, Technical University of Denmark, Building 352, 2800 Lyngby, Denmark.
[2] GN ReSound Denmark, Lautrupbjerg 9, 2750 Ballerup, Denmark.

Abstract *– This paper presents a two-stage approach for single-channel separation of dependent audio sources. The proposed algorithm is developed in the Bayesian framework and designed for general audio signals. In the first stage of the algorithm, the joint distribution of discrete Fourier transform (DFT) coefficients of the dependent sources is modeled by complex Gaussian mixture models in the frequency domain from samples of individual sources to capture the properties of the sources and their correlation. During the second stage, the mixture is separated through a generalized Wiener filter, which takes correlation term and local stationarity into account. The performance of the algorithm is tested on real audio signals. The results show that the proposed algorithm works very well when the dependent sources have comparable variances and linear correlation.*

Keywords *– monaural source separation, complex Gaussian mixture model, Gaussian statistical model of DFT coefficients*

I. INTRODUCTION

Source separation problem arises in a variety of signal processing applications. It can be categorized in several ways: Depending on the amount of available information about the mixing process and sources, it can be divided into blind source separation (BSS) and semi-BSS; According to the relation of n (the number of sources) and m (the number of sensors), it falls into the categories of an under-determined problem ($m < n$), even-determined problem ($m = n$) and over-determined problem ($m > n$); Based on the relation between sources, it is either a problem with independent sources or a problem with dependent sources.

Most of the source separation algorithms are based on the assumption that the sources are statistically independent, which holds in most cases. However, in some special audio applications such as feedback cancellation in hearing aids, the mixture contains dependent sources. The very few algorithms dealing with dependent sources include both semi-BSS techniques and BSS techniques. The semi-BSS techniques are heavily dependent on the nature of the problem and thus very ad-hoc. A typical example is presented in [1], where the structure of the mixing matrix and source covariance matrices are known beforehand. Instead of applying strong prior knowledge, the BSS techniques usually make strong assumptions on the properties of the sources [2][3][4], such as time-frequency sparsity, to solve the dependent source separation problem. When the sources are only linearly correlated through room impulse responses, the problem reduces to a convolutive BSS problem, which is a very active and challenging research area.

Compared with even- and over-determined problems, under-determined source separation problems are generally much more difficult due to the lack of constraints. Additional constraints are normally applied by making strong assumptions on the source characteristics, incorporating sources models or providing prior knowledge on the mixing process and/or signals. One powerful assumption about the sources is that they have a parsimonious representation in a given basis, such as the time-frequency (T-F) representation. This kind of assumption has lead to encouraging techniques [4][5][6]. Another class of methods incorporate source models, such as Vector Quantization (VQ), Gaussian Mixture Models (GMM), train the models first and separate the mixture afterwards based on proper criteria (e.g., minimum mean square error, likelihood ratio) [7][8].

The problem discussed in this paper, monaural separation of two general audio signals that are strongly dependent, is a combination of an under-determined problem and a problem with correlated sources. Most of the algorithms reviewed above fail in this extreme case either because the sparse representation for correlated sources is not valid or because the algorithms require multiple channels. For example, the existing technique for under-determined convolutive BSS requires multiple channels [9].

Since the sources are general, specific source models, such as speech model, are not applicable. Besides, how the two signals are correlated and what properties the sources exhibit are unknown. To combat these difficulties, a two-stage algorithm based on generalized Wiener filtering is proposed in this paper. In the first step, complex GMM is exploited to acquire sufficient knowledge about the sources and the way they are correlated. Based on the information obtained in the first step, the mixture is separated later by a generalized Wiener filter.

Although the study here is for two dependent sources, the method proposed can be generalized to more sources at least theoretically.

II. PROBLEM FORMULATION

The microphone signal x is a mixture of two dependent signals s_1 and s_2, i.e.,

978-1-4244-1835-0/07/$25.00 ©2007 IEEE

$$x = s_1 + s_2 \tag{1}$$

In the Bayesian framework, the two sources can be estimated through estimators such as maximum likelihood (ML). However, since the problem is under-determined, there will be multiple solutions with the ML estimator [7]. One of the alternatives is the maximum *a posteriori* (MAP) estimator:

$$\begin{aligned} (\hat{s}_1, \hat{s}_2) &= \arg\max_{s_1, s_2} p(s_1, s_2 \,|\, x) \\ p(s_1, s_2 \,|\, x) &\propto p(x \,|\, s_1, s_2)\, p(s_1, s_2) \end{aligned} \tag{2}$$

where $p(x|s_1,s_2)$ is the likelihood function, $p(s_1,s_2)$ is the prior knowledge about the joint distribution of the sources, which essentially reflects the statistical properties of each individual source and the correlation of the two sources.

A similar estimator is the conditional posterior mean (PM):

$$(\hat{s}_1, \hat{s}_2) = E\left[s_1, s_2 \,|\, x \right] \tag{3}$$

where the expectation operator $E[\bullet]$ implicitly requires the knowledge of the joint distribution $p(s_1,s_2)$.

Therefore, in the Bayesian framework, the solution for this source separation problem includes a stage of estimating the joint distribution and a second stage of separating the mixture based on a proper estimator such as MAP or PM.

III. ESTIMATION OF JOINT DISTRIBUTION

The estimation of the joint distribution can be performed in the time domain or any domain spanned by proper basis functions. Since the correlation between the two dependent sources usually varies strongly with frequency, time-domain modeling lacks the resolution to describe the difference among frequency bins and consequently leads to degraded performance. For discrete signals, discrete Fourier basis has several desirable properties and serves as an efficient domain for analyzing the signals in this paper.

A. Gaussian Statistical Model of DFT Coefficients

For real-time processing purpose, the sampled microphone signal $x(n)$ is broken into frames. Each frame is Fourier transformed. This process is referred to as short-time Fourier transform (STFT), i.e.,

$$\begin{aligned} &\tilde{X}_k(m) = \sum_{n=0}^{L-1} x((m-1)(L-M)+n+1)h(n)e^{-j\frac{2\pi}{L}kn} \\ &k = 0,1,\cdots,L-1 \qquad m = 1,2,\cdots \end{aligned} \tag{4}$$

where L is the length of each frame, M is the length of overlapping, m is the frame index, $h(n)$ is the window function applied. $\tilde{X}_0(m)$ and $\tilde{X}_{L/2}(m)$ are not interesting since they are direct current (DC) and Nyquist components respectively. $\tilde{X}_k(m)$ $k = L/2+1,\cdots,L-1$ are also ignored due to the symmetry of DFT coefficients. A tilde is used to denote a complex quantity. For brevity, m is dropped out in the following formulas, and $h(n)$ is also neglected.

A widely accepted assumption for stationary audio signals is that the DFT coefficients are statistically independent Gaussian random variables [11], i.e.,

$$\begin{aligned} &\begin{bmatrix} \tilde{X}_1 \\ \vdots \\ \tilde{X}_{\frac{L}{2}-1} \end{bmatrix} \sim CN(\boldsymbol{0}, diag\{\sigma_1^2, \cdots, \sigma_{\frac{L}{2}-1}^2\}) \\ &\sigma_k^2 = Var(\tilde{X}_{k,re}) + Var(\tilde{X}_{k,im}) \end{aligned} \tag{5}$$

where symbols with bold and italic font represent matrices or vectors, $diag\{\bullet\}$ is the diagonal matrix formed by the listed entries, $Var(\bullet)$ is the variance of the listed entries, subscripts 're' and 'im' denote the real part and imaginary part of a complex quantity respectively, and *CN* denotes the complex Gaussian distribution [10].

(5) implies that the DFT coefficients in different frequency bins are independent. It also implies that the real and imaginary parts of coefficients in each frequency bin are independent, have Gaussian distributions and the same variances. In a strict sense, the DFT coefficients follow an asymptotical Gaussian distribution as L approaches infinity [11].

To reduce the large number of parameters to estimate, it is assumed that the two correlated zero-mean stationary signals $s_1(n)$ and $s_2(n)$ are only correlated within the same frequency bin as shown in (6). This assumption usually holds very well for many types of correlation, especially linear correlation.

$$\begin{aligned} &\tilde{S}_{1,k_1,2,k_2} \sim CN(\boldsymbol{0}, \begin{bmatrix} \sigma_{1,k_1}^2 & 0 \\ 0 & \sigma_{2,k_2}^2 \end{bmatrix}) \qquad k_1 \neq k_2 \\ &\tilde{S}_{1,k,2,k} \sim CN(\boldsymbol{0}, \begin{bmatrix} \sigma_{1,k}^2 & \sigma_{12,k} \\ \sigma_{12,k}^* & \sigma_{2,k}^2 \end{bmatrix}) \qquad \sigma_{12,k} = E[\tilde{S}_{1,k}^* \tilde{S}_{2,k}] \\ &\tilde{S}_{1,k_1,2,k_2} = [\tilde{S}_{1,k_1}, \tilde{S}_{2,k_2}]^T \\ &k, k_1, k_2 = 1,\cdots, L/2-1 \end{aligned} \tag{6}$$

To obtain an expressible probability density function in terms of $\tilde{S}_{1,k_1,2,k_2}$, we have to further assume a special relation between the covariance matrices of $\tilde{S}_{1,k_1}$ and $\tilde{S}_{2,k_2}$ [12]:

$$\begin{aligned} &Cov(\tilde{S}_{1,k_1,re}, \tilde{S}_{2,k_2,re}) = Cov(\tilde{S}_{1,k_1,im}, \tilde{S}_{2,k_2,im}) \\ &Cov(\tilde{S}_{1,k_1,re}, \tilde{S}_{2,k_2,im}) = -Cov(\tilde{S}_{2,k_2,re}, \tilde{S}_{1,k_1,im}) \\ &k_1, k_2 = 1,\cdots, L/2-1 \end{aligned} \tag{7}$$

where $Cov(\bullet)$ is the covariance of the listed entries. It was found that the assumption (7) holds well at least for linear correlation.

B. GMM Estimation of the Joint Distribution

As seen above, the DFT coefficients of a stationary audio signal can be described by Gaussian distributions. Therefore, its power spectral density, which gives the variance as a

function of frequency, is completely taken into account by the Gaussian distributions. However, realistic audio signals are only locally stationary and contain various types of timbres and pitches [13]. The complex GMM, instead of a single complex Gaussian distribution, has to be adopted to capture the diverse spectra of the signals. As a semi-parametric method to estimate the probability density function, GMM also possesses the advantages of high flexibility and reasonable complexity compared with non-parametric and parametric methods [14].

In each frequency bin, the joint distribution of DFT coefficients of the two dependent zero-mean signals is modeled by the complex GMM as below:

$$p(\tilde{S}_{1,k,2,k}) = \sum_{i=1}^{Q} \omega_{i,k}\, p_G(\tilde{S}_{1,k,2,k}, \boldsymbol{C}_{i,k})$$
$$\boldsymbol{C}_{i,k} = \begin{bmatrix} \sigma^2_{1,i,k} & \sigma_{12,i,k} \\ \sigma^*_{12,i,k} & \sigma^2_{2,i,k} \end{bmatrix}, \sum_{i=1}^{Q} \omega_{i,k} = 1 \qquad (8)$$
$$k = 1, \cdots, L/2 - 1$$

where Q is the number of components in GMM, i is the index to the ith Gaussian components, $p_G(\tilde{S}_{1,k,2,k}, \boldsymbol{C}_{i,k})$ is the centered complex Gaussian distribution with covariance matrix $\boldsymbol{C}_{i,k}$, $\omega_{i,k}$ is the weight of the ith Gaussian component in the kth frequency bin.

By fixing the mean of each component as zero, the number of parameters to be estimated in each frequency bin is further reduced to $4Q$, including $3Q$ in the covariance matrix $\boldsymbol{C}_i(k)$ and Q in the weights ω_i.

The $4Q$ parameters are estimated in the first stage of the algorithm from samples of individual sources by standard expectation-maximization (EM) algorithm with K-means initialization. On-line EM algorithm can also be applied to enable a real-time implementation [15].

IV. SEPARATION OF MIXTURE

In the separation stage, the traditional Wiener filter [16] is extended in two aspects to separate the mixture of the two dependent sources. Firstly, it is generalized to take the correlation between the dependent sources into consideration. Secondly, the fixed gain of the Wiener filter is extended to be adaptive so that the local stationarity can be dealt with. These two aspects lead to the design of an adaptive weighted Wiener filter.

Based on the information obtained in the first stage, the two sources can be estimated through the MAP estimator in each frequency bin:

$$(\hat{\tilde{S}}_{1,k}, \hat{\tilde{S}}_{2,k}) = \arg\max_{\tilde{S}_{1,k}, \tilde{S}_{2,k}} p(\tilde{S}_{1,k}, \tilde{S}_{2,k} \,|\, \tilde{X}_k) \qquad (9)$$

However, (9) is not directly tractable [7]. To get back to the traditional Wiener filtering case, a hidden random variable q is introduced, which is associated with the active Gaussian component in GMM and is referred to as state variable. The posterior probability in (9) is then formulated as:

$$p(\tilde{S}_{1,k}, \tilde{S}_{2,k} \,|\, \tilde{X}_k) = \sum_{j=1}^{Q} p(\tilde{S}_{1,k}, \tilde{S}_{2,k} \,|\, \tilde{X}_k, q = j) p(q = j \,|\, \tilde{X}_k) \qquad (10)$$

Therefore, the estimation of sources needs three steps: estimate the current state by calculating the posterior probability of the state variable; construct the filters by maximizing the posterior probability of the sources given the state; separate the mixture and reconstruct the two sources in the time domain. In the following formula, $q = j$ is abbreviated as q_j.

A. State Estimation

The state variable q can be estimated through the posterior probability, i.e., $p(q_j \,|\, \tilde{X}_k)$, denoted as $\gamma_{j,k}$. It is calculated as:

$$\gamma_{j,k} \propto p(\tilde{X}_k \,|\, q_j) p(q_j) = p(\tilde{X}_k \,|\, q_j) \omega_{j,k} \qquad (11)$$

When the active state is given as j, $\tilde{X}_k$ is the sum of two correlated complex Gaussian variables with the joint distribution:

$$\left[\tilde{S}_{1,k}, \tilde{S}_{2,k}\right]^T \sim CN(\boldsymbol{\theta}, \boldsymbol{C}_{j,k}) \qquad (12)$$

where $\boldsymbol{C}_{j,k}$ is given in (8). It can be shown $\tilde{X}_k$ follows:

$$p(\tilde{X}_k \,|\, q_j) = p_G(\tilde{X}_k, \sigma^2_{1,j,k} + \sigma^2_{2,j,k} + 2\sigma_{12,j,k}) \qquad (13)$$

Inserting (13) into (11), we obtain,

$$\gamma_{j,k} \propto \omega_{j,k}\, p_G(\tilde{X}_k, \sigma^2_{1,j,k} + \sigma^2_{2,j,k} + 2\sigma_{12,j,k}) \qquad (14)$$

B. Construction of the filters

Given the active state q, (9) can be solved by extending the Wiener filter.

It is obvious that

$$p(\tilde{X}_k \,|\, \tilde{S}_{1,k}, \tilde{S}_{2,k}, q_j) = \delta(\tilde{S}_{1,k} + \tilde{S}_{2,k} - \tilde{X}_k) \qquad (15)$$

where $\delta(\bullet)$ is the Dirac delta function.

$p(\tilde{S}_{1,k}, \tilde{S}_{2,k} \,|\, q_j)$, the likelihood of the hidden q process can be calculated straightforward:

$$p(\tilde{S}_{1,k}, \tilde{S}_{2,k} \,|\, q_j) = p_G(\tilde{S}_{1,k,2,k}, \boldsymbol{C}_{j,k}) \qquad (16)$$

Therefore, given the active component in GMM, the posterior probability of the two sources is:

$$p(\tilde{S}_{1,k}, \tilde{S}_{2,k} \,|\, \tilde{X}_k, q_j) \propto p(\tilde{X}_k \,|\, \tilde{S}_{1,k}, \tilde{S}_{2,k}, q_j) p(\tilde{S}_{1,k}, \tilde{S}_{2,k} \,|\, q_j)$$
$$= \delta(\tilde{S}_{1,k} + \tilde{S}_{2,k} - \tilde{X}_k)\, p_G(\tilde{S}_{1,k,2,k}, \boldsymbol{C}_{j,k}) \qquad (17)$$

The MAP estimator (9)-(10) can be solved by picking up the Gaussian component with the highest probability calculated in (14) and maximizing (17) under the constraint:

$$\tilde{S}_{1,k} + \tilde{S}_{2,k} = \tilde{X}_k \qquad (18)$$

The solution can be easily found as:

$$\hat{\tilde{S}}_{1,k} = \frac{\sigma_{1,j,k}^2 + \mathrm{Re}(\sigma_{12,j,k})}{\sigma_{1,j,k}^2 + \sigma_{2,j,k}^2 + 2\,\mathrm{Re}(\sigma_{12,j,k})}\tilde{X}_k$$
$$\hat{\tilde{S}}_{2,k} = \frac{\sigma_{2,j,k}^2 + \mathrm{Re}(\sigma_{12,j,k})}{\sigma_{1,j,k}^2 + \sigma_{2,j,k}^2 + 2\,\mathrm{Re}(\sigma_{12,j,k})}\tilde{X}_k \quad (19)$$

where j corresponds to the component with highest $\gamma_{j,k}$ in the kth frequency bin.

An alternative estimator PM in (3) assigns every Gaussian component with a probability instead of a hard decision on active components, which is shown below:

$$\begin{aligned} E[\tilde{S}_{1,k}\,|\,\tilde{X}_k] &= \int_{\tilde{S}_{1,k}} (\tilde{S}_{1,k} \int_{\tilde{S}_{2,k}} p(\tilde{S}_{1,k}, \tilde{S}_{2,k}\,|\,\tilde{X}_k) d\tilde{S}_{2,k}) d\tilde{S}_{1,k} = \\ &\int_{\tilde{S}_{1,k}} (\tilde{S}_{1,k} \int_{\tilde{S}_{2,k}} (\sum_{j=1}^{Q} p(\tilde{S}_{1,k}, \tilde{S}_{2,k}\,|\,\tilde{X}_k, q_j) p(q_j\,|\,\tilde{X}_k)) d\tilde{S}_{2,k}) d\tilde{S}_{1,k} \\ &= \sum_{j=1}^{Q} \gamma_{j,k} (\int_{\tilde{S}_{1,k}} (\tilde{S}_{1,k} \int_{\tilde{S}_{2,k}} p(\tilde{S}_{1,k}, \tilde{S}_{2,k}\,|\,\tilde{X}_k, q_j) d\tilde{S}_{2,k}) d\tilde{S}_{1,k}) \\ &= \sum_{j=1}^{Q} \gamma_{j,k} E\left[\tilde{S}_{1,k}\,|\,\tilde{X}_k, q_j\right] \end{aligned} \quad (20)$$

Since $\tilde{S}_{1,k}$ follows a Gaussian distribution for a given active state q, its mean corresponds to the peak location. In other words, the MAP estimation (19) can replace $E\left[\tilde{S}_{1,k}\,|\,\tilde{X}_k, q_j\right]$ in (20). Therefore the PM estimation for the two sources is:

$$\hat{\tilde{S}}_{1,k} = \sum_{j=1}^{Q} \gamma_{j,k} \frac{\sigma_{1,j,k}^2 + \mathrm{Re}(\sigma_{12,j,k})}{\sigma_{1,j,k}^2 + \sigma_{2,j,k}^2 + 2\,\mathrm{Re}(\sigma_{12,j,k})}\tilde{X}_k$$
$$\hat{\tilde{S}}_{2,k} = \sum_{j=1}^{Q} \gamma_{j,k} \frac{\sigma_{2,j,k}^2 + \mathrm{Re}(\sigma_{12,j,k})}{\sigma_{1,j,k}^2 + \sigma_{2,j,k}^2 + 2\,\mathrm{Re}(\sigma_{12,j,k})}\tilde{X}_k \quad (21)$$

(21) can be regarded as a generalized Wiener filter. It separates the mixture by considering the correlation term between the signals. Besides, it is weighted by the posterior probability $\gamma_{j,k}$. Since $\gamma_{j,k}$ is adaptive for locally stationary signals, (21) is essentially a weighted adaptive Wiener filter.

V. SIMULATION RESULTS

The proposed algorithm is evaluated on a male speech s_1, which is filtered by a 128-tap impulse response shown in Figure 1 to form a linearly correlated source s_2. The impulse response has a shape of typical response of feedback path. It is chosen so that the correlation between s_1 and s_2 varies with frequencies, and the variances of s_1 and s_2 are comparable. The first 45 seconds of the two signals are used to train the GMM. The following 15 seconds are mixed for separation.

The PM estimator in (21) is selected for the simulation since in GMM it is usually superior to the MAP estimator as shown in [7]. The frame length is 512 samples, corresponding to approximately 5 milliseconds. The number of Gaussian components is 3.

The popular measures of source separation performance, such as in [17], usually apply for independent sources. For dependent sources, the measure adopted in this paper is a simple normalized test error, defined in the time domain as:

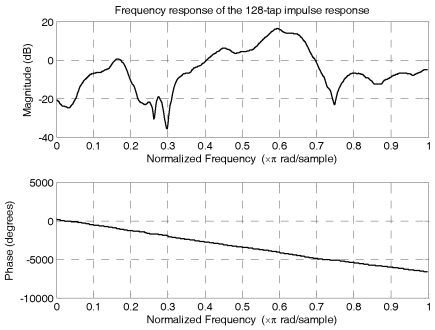


Fig. 1. Frequency response of the 128-tap echo-like impulse response

$$\varsigma_i = \frac{\left\|\hat{s}_i - s_i\right\|_2}{\left\|s_i\right\|_2}, i = 1, 2 \quad (22)$$

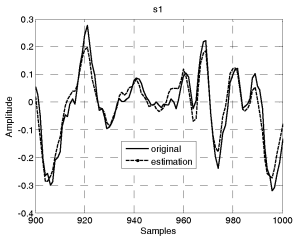
where $\|\cdot\|_2$ denotes the L2-norm.

The separation results are shown in Figure 2. Figure 2(a) and 2(b) illustrate the partial waveforms of the two sources. Figure 2(c) and 2(d) are original spectrograms of the two signals. Figure 2(e) and 2(f) are the estimated spectrograms. The original speech signal is cut off at 4 kHz. Although the estimated signal is not sharply cut off there due to a higher sampling rate (11025 Hz), the estimated frequency contents are small enough above 4 kHz. The comparison between the performance of traditional Wiener filter and the proposed algorithm is given in Table 1. Figure 2 and Table 1 show that the proposed algorithm can separate the excerpted signals very well.

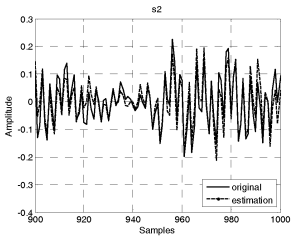
It is also noted that there exists a pattern in Figure 2(e) and 2(f): Horizontal broken lines are located at some evenly spaced frequency bins. This indicates the failure of modeling at these frequencies. One possible reason is that voiced speech shows strong tonal characteristics at harmonic frequencies. The DFT coefficients at these frequencies tend to be constant, which GMM with zero-mean components is not able to model. The other possible reason is that the proposed algorithm is phase blind, which is inherited from Wiener filtering. In some frequency bins, the two sources could be negatively correlated. The mixture is therefore the remaining signal after mutual cancellation. The amount of cancellation is impossible to recover when the phase information is missing. The separation performance is thus severely degraded in those frequency bins.

Table 1. Comparison between the performance of Wiener filter and generalized Wiener filter (three Gaussian components)

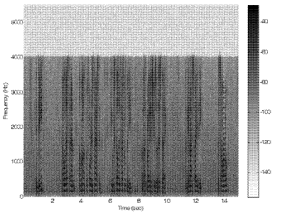
Normalized Error	Source 1	Source 2
Wiener Filter	0.4060	0.4351
Generalized Wiener Filter	0.3497	0.3742



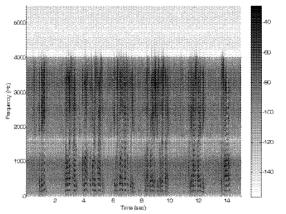
(a) partial waveform of source 1



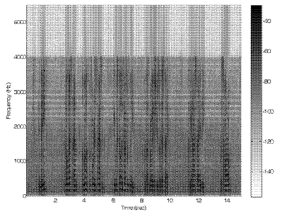
(b) partial waveform of source 2



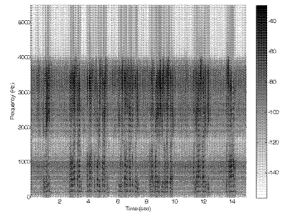
(c) original spectrogram of source 1



(d) original spectrogram of source 2



(e) estimated spectrogram of source1



(f) estimated spectrogram of source2

Fig. 2. Simulation results with 3 Gaussian components and 512-sample frame length ($\varsigma_1 = 0.3497, \varsigma_2 = 0.3742$)

The separation performance as a function of the frame size and the number of Gaussian components is studied. In Figure 3, as the number of Gaussian components increases, the normalized test error drops down first and increases later. The drop indicates the inadequacy of Gaussian components in describing the two locally stationary dependent signals. The following rise is probably a consequence of over-fitting. The optimal number for the excerpted signals is three. In Figure 4, it is seen that the performance is generally better with a finer spectrum, i.e., longer frames. Since time domain modeling can be regarded as a special frequency-domain modeling when the frame length is equal to 1, this infers that time-domain modeling is worse than frequency-domain modeling.

It can be noted that the test error of s_2 is consistently higher than s_1 . This is because s_1 has a larger variance than s_2 . Constraint (18) requires:

$$x(n) = s_1(n) + s_2(n) = \hat{s}_1(n) + \hat{s}_2(n)$$
$$s_1(n) - \hat{s}_1(n) = s_2(n) - \hat{s}_2(n) \quad (23)$$

Therefore,

$$\frac{\|s_1 - \hat{s}_1\|_2}{\|s_1\|_2} < \frac{\|s_2 - \hat{s}_2\|_2}{\|s_2\|_2}, if \ \|s_1\|_2 > \|s_2\|_2 \quad (24)$$

i.e.,

$$\varsigma_1 < \varsigma_2 \quad (25)$$

Thus the dominant source is always estimated better. The more dominant the better estimated.

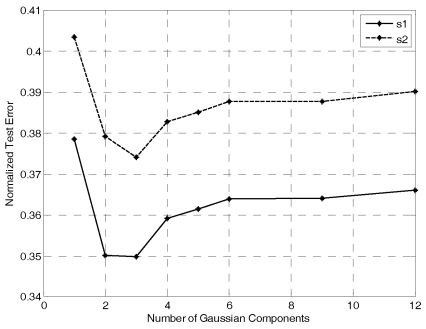


Fig. 3. Normalized test error as a function of the number of Gaussian components

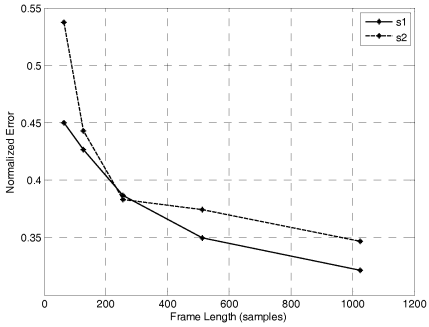


Fig. 4. Normalized test error as a function of frame length

VI. CONCLUSIONS AND FUTURE WORK

A two-stage algorithm for monaural separation of dependent audio sources is proposed in this paper. The algorithm is based on the Bayesian framework and designed for general audio sources, not restricted to speech or music. In the first stage of the algorithm, complex GMM is used in the frequency domain to capture the properties of the signals and their correlation. In the second stage, based on the information obtained, traditional Wiener filter is extended to an adaptive weighted Wiener filter to separate the mixture. The extension takes both correlation term and local stationarity into account. The simulation results indicate that the proposed algorithm performs very well when the two sources are linearly correlated and have comparable variances.

As explained in (23)-(25), when the two sources have incomparable variances, i.e. one of the sources dominates, the stronger source is always estimated better. This is an inherent property of the measure and constraint (18). It could also be regarded as one of the limitations in many source separation algorithms including the proposed algorithm in this paper.

Due to the loss of phase information in Wiener filtering, when the two sources are negatively correlated, it is very hard to estimate the amount of cancellation in the mixture for an under-determined source separation problem. An additional phase model could be added to the training step of the algorithm to obtain the phase information. A relatively easy way is to model the phase difference between the two signals in each frequency bin because it is usually more stable than the phases of individual sources, especially when the correlation of the two sources is time-invariant or slowly time-varying. A successful phase modeling should improve the performance of the algorithm significantly.

Another aspect of future work is to evaluate the algorithm with non-linearly correlated sources. This may lead to some modifications of the algorithm, such as modeling the correlation between adjoining frequency bins.

It should be noted that the proposed algorithm needs more signal samples to assess the performance before conclusions about the best choice of components and frame length are drawn.

REFERENCES

[1] L. Bedini, D. Herranz, E. Salerno, C. Baccigalupi, E. E. Kuruoglu, and A. Tonazzini, "Separation of Correlated Astrophysical Sources Using Multiple-Lag Data Covariance Matrices", *EURASIP Journal on Applied Signal Processing*, vol. 2005, Issue 15, pp. 2400-2412.

[2] Frederic Vrins, John A. Lee, and Michel Verleysen, "Filtering-Free Blind Separation of Correlated Images", *the 8th International Workshop on Artificial Neural Networks*, 2005.

[3] Cichocki A., and Georgiev P., "Blind Separation Algorithms with Matrix Constraints", *IEICE Transactions Fundamentals of Electronics Communications and Computer Science*, E86-A, pp. 522-531, 2003.

[4] F. Abrard, and Y. Deville, "A Time-frequency Blind Signal Separation Method Applicable to Underdetermined Mixtures of Dependent Sources", *Signal Processing*, vol. 85, pp. 1389-1403, July 2005.

[5] F. Abrard, Y. Deville, "Blind Separation of Dependent Sources Using the Time-frequency Ratio of Mixtures Approach", *the 7th International symposia Signal Processing Applications (ISSPA)*, July 2003.

[6] Yuanqing Li, Shun-Ichi Amari, Andrzej Cichocki, Daniel, W. C. Ho, and Shengli Xie, "Underdetermined Blind Source Separation Based on Sparse Representation", *IEEE Transactions on Signal Processing*, vol.54, no. 2, pp. 423-437, 2006.

[7] Benaroya, L., Bimbot, F., and Gribonval, R., "Audio Source Separation with a Single Sensor", *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 191-199, January 2006.

[8] Daniel P.W.Ellis, Ron J.Weiss, "Model-based Monaural Separation Using a Vector-Quantized Phase-Vocoder Representation", *ICASSP 2006*.

[9] Thomas Melia and Scott Rickard, "Underdetermined Blind Source Separation in Echoic Environments Using DESPRIT", *EURASIP Journal on advances in Signal Processing*, vol. 2007, Article ID 86484.

[10] Steven M. Kay, "Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory", *Prentice Hall PTR*, 1993.

[11] William A. Pearlman, and Robert M. Gray, "Source Coding of the Discrete Fourier Transform", *IEEE Transactions on Information Theory*, vol. 24, no. 6, pp. 683-692, November 1978.

[12] R. A. Wooding, "The Multivariate Distribution of Complex Normal Variables", *Biometrika*, vol. 43, no. 1-2, pp. 212-215, 1956.

[13] Laurent Benaroya, and Frederic Bimbot, "Wiener Based Separation with HMM/GMM Using a Single Sensor", *Proceedings 4th International Symposia on Independent Component Analysis and Blind Signal Separation*, April 2003.

[14] Christopher M. Bishop, "Neural Networks for Pattern Recognition", pp. 33, *Oxford University Press*, November 1995.

[15] Masa-aki Sato, and Shin Ishii, "On-line EM Algorithms for the Normalized Gaussian Network", *Neural Computation*, vol.12, no. 2, pp. 407-432, 2000.

[16] Wiener, Norbert, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series", *The MIT Press*, 1949.

[17] E. Vincent, R. Gribonval, and C. FEVOTTE, "Performance Measurement in Blind Audio Source Separation", *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1462-1469, July 2006.